DOCUMENT RESUME

ED 328 463 SE 051 948

AUTHOR Livingstone, Ian D.; And Others
TITLE The Computer as a Diagnostic Tool in Mathematics. Final Report--November 1988. Evaluation of Exploratory Studies in Educational Computing Study 13.
INSTITUTION New Zealand Council for Educational Research, Wellington.
REPORT NO ISBN-0-908567-77-4
PUB DATE 88
NOTE 90p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

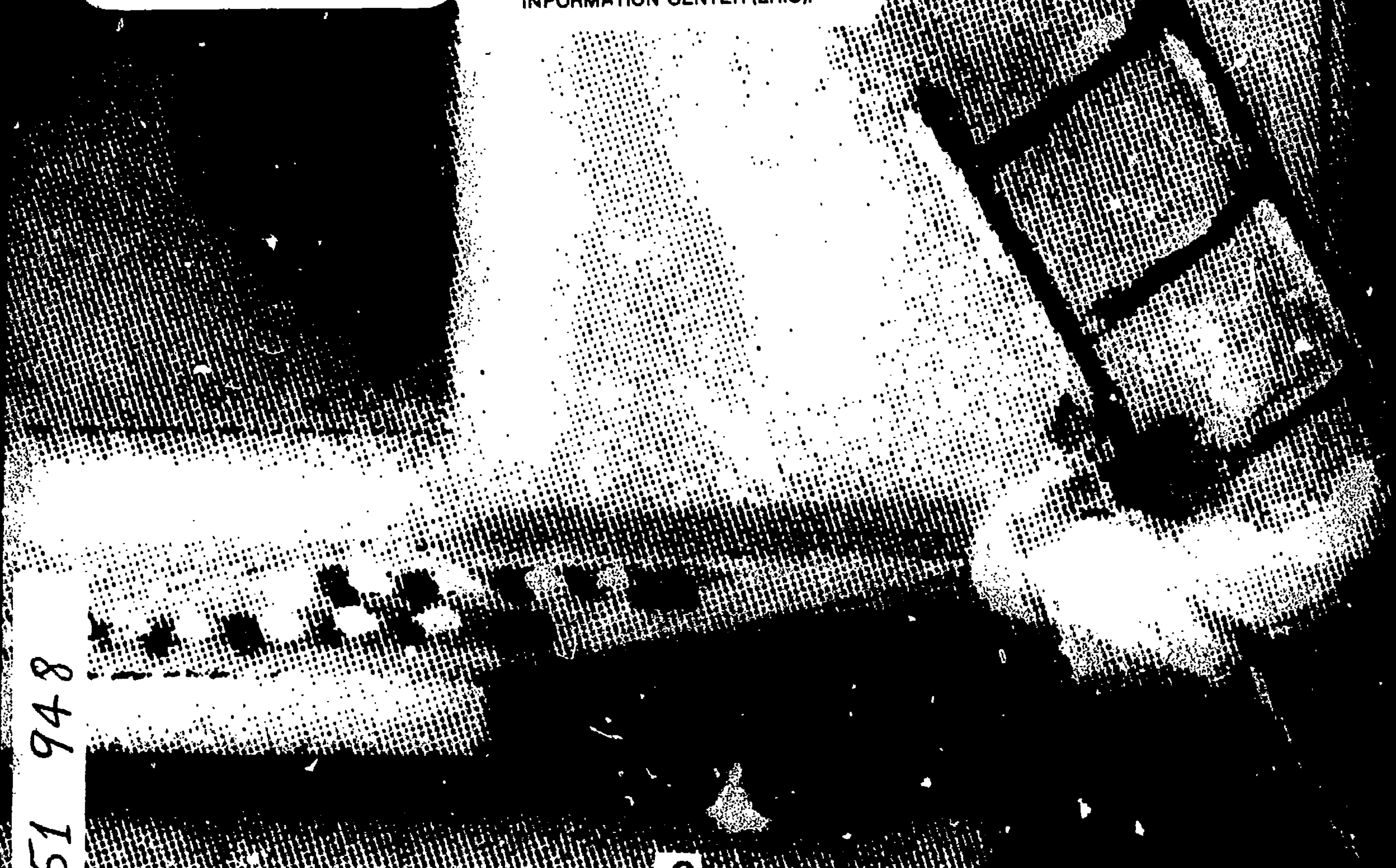
EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Arithmetic; *Computer Assisted Instruction; Computer Uses in Education; *Diagnostic Tests; Elementary Education; *Elementary School Mathematics; *Foreign Countries; *Mathematics Achievement; Mathematics Education; Problem Solving
IDENTIFIERS *New Zealand

ABSTRACT

The aim of this study was to investigate and validate the use of a computerized testing program for the diagnosis of arithmetic difficulties experienced by primary school children. The basic research question was whether a microcomputer could be used to diagnose difficulties in addition, subtraction and multiplication as well as a paper-and-pencil test can. Variables considered were convenience and ease of use, time taken, the accuracy of the diagnosis in comparison with that of an experienced remedial teacher and of a regular classroom teacher, and the usability of the information provided. An introduction, aim of evaluation, general procedure, analysis of results, formative assessment, and conclusions are provided. A trial version and revised version of the Seville Diagnostic Arithmetic Test, error scatterplots by tester, and error comparisons by levels are appended. (KR)

Evaluation of Exploratory Studies in Educational Computing

STUDY 13: THE COMPUTER AS A DIAGNOSTIC TOOL IN MATHEMATICS



2

New Zealand Council for Educational Research

Evaluation of
Exploratory Studies in Educational Computing

STUDY 13: THE COMPUTER AS A DIAGNOSTIC TOOL IN MATHEMATICS

FINAL REPORT – NOVEMBER 1988

Ian D. Livingstone
Barry Eagle
John Laurie

New Zealand Council for Educational Research
Wellington

New Zealand Council for
Educational Research
P.O. Box 3237
Wellington
New Zealand

ISBN 0-908567-77-4

ABSTRACT

The aim of this study was to investigate and validate the use of a computerized testing program for the diagnosis of arithmetic difficulties experienced by primary school children. The program, based around a modification of the Seville Diagnostic Arithmetic Test, had been designed and written by mathematics staff at Christchurch Teachers College, and was tested on children in Standards 2, 3 and 4 in three Christchurch schools during 1986 and 1987.

The basic research question was whether a microcomputer could be used to diagnose difficulties in addition, subtraction and multiplication as well as a paper-and-pencil test can. Variables considered were convenience and ease of use, time taken, the accuracy of the diagnosis in comparison with that of an experienced remedial teacher and of a regular classroom teacher, and the usability of the information provided.

Computers were introduced into the test classrooms with language development and 'educational game' type software to familiarise students with their presence and operation. During the test phase, children worked through computer-generated and randomly presented arithmetic examples of increasing complexity. Selected children were also tested in the traditional way by their classroom teacher or an itinerant remedial teacher on a modified version of the Seville Diagnostic Arithmetic Test, and their results were compared with those obtained from the computer.

The results were encouraging. The study showed that the software (which underwent continuous revision during the study) was able to diagnose areas of misunderstanding in basic arithmetic operations quite successfully and efficiently, allowing teachers to concentrate teaching on points of specific difficulty. Children at the Standard 2 level found difficulties in coping with the computer diagnostic program, but Standard 3 and 4 children handled it well, and showed very positive attitudes to the experience.

TABLE OF CONTENTS

TABLES

PREFACE

The Exploratory Studies in Educational Computing (ESEC) were set up at the request of, and funded by, the New Zealand Minister of Education, following a new policy provision introduced in 1985. The purpose of the studies was to provide a basis for future policy developments in educational computing.

Initial proposals were sought in an advertisement in the *Education Gazette* of 14 June, 1985, and some 200 separate proposals from more than 100 schools were received. A broadly representative conference met at the Stella Maris Retreat Centre, Wellington, between 2-6 September, 1985 to consider the applications, and eventually 15 distinct studies for major funding were chosen. Subsequently two of these were subdivided, making 19 separate studies in all.

The Computers in Education Development Unit (CEDU), within the Department of Education, was responsible for the technical management and funding of the projects, and with the exception of one study, the New Zealand Council for Educational Research has been responsible for their evaluation. (One is being evaluated at the University of Auckland.) Each study was co-ordinated and conducted by a committee consisting of the teachers involved (who were often, though not always, the originators of the study), one member from the CEDU, at least one member from the NZCER, and often others from the inspectorate, teachers colleges or regional resource centres.

Many of the proposals had requested specific computer equipment and software, and this was ordered and shipped to schools by the beginning of 1986. Classroom computer work commenced at various times during 1986, and proceeded through 1987. Various research materials were prepared for use as required by all the studies, including pre- and post-questionnaires for students and teachers, and logs and diaries to record day-to-day impressions. In addition, study-specific instruments were prepared where necessary. One of the studies was at the preschool level, and four studies dealt with children with special needs. All of the remainder were located in primary schools, but some involved secondary school children as well.

The projects are distinctive in the way in which they have been initiated by classroom teachers, rather than by Departmental policy-makers or educational reseachers. The level of commitment from all the teachers involved in the projects has consequently been very high. They responded positively to the opportunity to participate, and contributed many hours of extra work to the evaluative aspects of the studies.

The study reported in this report, Study No. 13, involved teachers from three Christchurch schools: Somerfield Contributing School, Redcliffs Primary School, and Elmwood Normal School. To these teachers, their principals, and all the Standard 2, Standard 3 and Standard 4 children who took part in the experiment and its evaluation go our warmest thanks. We hope you got as much out of the experience as you gave to it.

Ian D. Livingstone
Barry Eagle
John Laurie

November 1988

1 INTRODUCTION

Behind most approaches to educational diagnosis has been the use of tests to provide information about specific problems in the performance of a task by an individual student, information which it is hoped will point to some form of appropriate remedial treatment. To arrive at such a diagnosis, these tests generally are concerned with the following key elements:

a) examination of a student's consistent errors;
b) construction of a profile of a student's strengths and weaknesses;
c) identification of the specific misunderstandings which have led the student to perform poorly.

Early this century, Anderson (1918) discussed diagnostic testing with reference to seven types of errors in long division, with the aim of enabling teachers to diagnose what he termed 'mathematical diseases'. Thus, right from the very beginning, a 'medical' model has been applied to educational diagnosis, on the assumption that if a particular pattern of errors can be detected, then an appropriate remedy can be 'prescribed'. However, it is by no means certain that this is the case, and diagnostic testing in education has been criticised as building on weak theoretical foundations. In the past, it has generally been tied to paper-and-pencil tests, often administered by a classroom teacher, sometimes by a specialist. Close interrogation of students as they solve problems has been emphasised, but this is really only practical in a clinical, one-to-one situation. Thus, despite good intentions, progress has been slow in developing useful diagnostic tests which could provide the busy classroom teacher with a profile of the specific errors made by students, even in relatively straightforward subjects like mathematics.

Basic to a desirable approach to diagnosis is the idea of a *pattern* of performance, requiring an understanding of the actual *nature* of the erroneous responses given by students, and not simply the *number* incorrect. As it is the incorrect answers which provide diagnostic clues about the difficulties which a student is finding, ways to classify and analyse these responses are needed before detailed remediation is possible. The alternative, of course,

is simply to re-teach the material related to the particular objectives on which students are failing, until mastery is eventually obtained. There are some advocates of this 'broad brush' approach as being the most practical in a typical classroom situation.

Thomas (1981) provides a useful diagnostic model containing three stages: *Status Assessment*, which asks critical questions about the specific objectives the student is expected to have achieved, what assessment techniques can best determine how well the student has achieved those objectives, and what pattern of discrepancies between expectations and performance is identified by those techniques; *Cause Estimation*, which asks what reasons for the deficiencies need to be considered, how can these possibilities be evaluated, and what is the most likely cause (or combination of causes) for the pattern of errors found; and *Treatment*, which enquires, in the light of the above, what treatments would help the student most effectively, what evaluation techniques are available to determine how well the treatment is succeeding, and how successful it is.

Over the past two decades, research has proceeded in a number of countries, in response to an increased awareness of a need for reliable diagnostic information in mathematics, if individual needs are going to be met adequately. One example is the extensive work by Hart (1981) and others at Chelsea College, University of London. She found a very wide spread of attainment, a 'seven year gap', between the levels of performance of secondary school pupils in the same year. A common mathematical 'diet' cannot hope to cater for such a spread. Bennett *et al* (1984) found that there was a poor match between the number tasks that primary school pupils were set and their grasp of number; many of the high achievers were set tasks that were too easy, while the low achievers were given work that was too hard.

Denvir and Brown (1987) examined the feasibility of a class-administered diagnostic test in primary mathematics, and came to the conclusion that the instrument they had prepared was able to provide an initial assessment of pupils' understanding of number from which those needing further diagnostic assessment by interview could be identified.

Adaptive testing, in which the sequence of items presented to a student depends on the student's previous response, really did not come into its own until the advent of the microcomputer in schools, but an ambitious early attempt in the United States led to the development of the *KeyMath Diagnostic Arithmetic Test* (Connolly, Nachtman, and Patchett, 1971). In this test, the diagnostic profile is developed on a large sheet of paper, which provides a map of arithmetic attainment, with the different content areas listed down the

page, and item difficulty levels moving from 'easy' on the left to 'difficult' on the right. Scaling according to the Rasch latent-trait model was used to establish the relative difficulties of the items.

Numerous attempts have been made in recent years to exploit computer technology to make use of the information contained in incorrect answers. Brown and Burton (1978) developed 'BUGGY', a computerized system for training teachers in diagnostic skills. The computer takes the role of the student responding to addition and subtraction questions and, by showing how test questions are answered incorrectly by the application of a particular incorrect rule, trains teachers to recognize the probable sources of error. Under this system, as long as the various 'bugs' are independent of one another, errors can be diagnosed easily.

Extensive work on the classification of errors at the Computer-based Education Research Laboratory at the University of Illinois, Champaign-Urbana (e.g., Tatsuoka and Birembaum, 1981), has concentrated on the skills of addition and subtraction. A major concern of this group was that students might obtain the right answer to a question by applying the wrong reasoning. Using a set of arithmetic test questions which incorporated up to 45 separate erroneous rules, these investigators showed that it is possible to infer from responses to related items whether a student has used an incorrect rule to obtain the correct answer to an item.

However, busy classroom teachers are unlikely to be able to handle such detailed and complex forms of diagnosis, and a more general approach to diagnostic assessment seems desirable. Rather than attempt an extended logical analysis of *possible* errors, a more practical approach would be to catalogue the *actual* errors that students make, and then write multiple choice items in which the incorrect alternatives (distracters) reflect these misconceptions. However, the use of multiple-choice items, rather than open-ended questions, does mean that the student making an unusual, idiosyncratic response needs a 'None of these' category in which to write a complete answer, if valuable information is not to be lost. This approach was used in devising the current series of Topic Pre-tests in Mathematics, designed for Form 3 students, and now being prepared by the New Zealand Council for Educational Research, in collaboration with the Department of Education and The Auckland College of Education. A particular feature of these tests is the novel, carbon-backed answer sheet which provides some information to allow pupils themselves to diagnose their own likely errors, and directs them to appropriate work-sheets, thus freeing the classroom teacher to concentrate on pupils with major or idiosyncratic difficulties.

Efforts to develop accurate and efficient computer-based diagnostic testing procedures during the last decade have met with mixed success. A relatively simple approach was employed in *Diagnose*, a computer-based program for reporting criterion-referenced test results (Furlong & Miller, 1978). It showed questions answered incorrectly, and provided a list of course objectives in need of further study, together with summary profiles of class performance for the teacher to study. But it was discontinued through a combination of lack of promotion and insufficient demand. Other promising schemes elsewhere in the world have foundered for similar reasons.

In Australia, the Tasmanian Education Department Diagnostic Information Service (TEDDIS) has recently released a comprehensive computer package of materials developed with the specific purpose of finding and treating errors in the ways in which students handle basic computations. The rationale behind the project was that it was no use finding out what a student was doing incorrectly unless there was also a firm intention of trying to correct the errors. Thus remediation, through the provision of a computer printout of an elaborate series of error codes associated with each incorrect answer, and indexed to appropriate teacher guides and student worksheets and activities materials, is an important part of the development. The tests cover the manipulation of both whole and rational numbers (Smith, 1987).

Current research efforts, particularly in the United States, have developed more thorough-going computer-diagnostic strategies, in which the students themselves respond to questions presented on a computer screen. This is a significant new development, and reflects the growing presence of the microcomputer in many school classrooms around the world. One such experiment has explored the nature of student misunderstanding, by applying the 'answer-until-correct' method, in which a student is shown the next item in sequence only after finding the correct response for a given item. This approach extracts a large amount of information about a student's ability from a given number of items, and goes some way towards distinguishing part mastery from complete mastery (Choppin, 1983). The present study, which uses the computer to generate at random a series of items of comparable difficulty, all testing the same objective, has some of the characteristics of this approach.

Another such example is *Math Doctor, M.D. - Microcomputer Adaptive Diagnosis* (Signer, 1982) a computer program designed to diagnose achievement in arithmetic number concepts, addition, subtraction, multiplication and division. This program uses a random generating function to construct examples, and branches to a new category when a student misses an objective which is a critical prerequisite skill for higher level items.

A more sophisticated testing strategy showing considerable promise is computerized adaptive testing (Weiss, 1983). The two features that distinguish this form of testing from its conventional counterpart are implied by its name: computerized test administration and adaptive test design. First, the examinee uses a standard keyboard or specially designed auxilliary device to answer questions that are displayed on the computer screen. Secondly, by making use of recent developments in Item Response Theory (IRT), the tests can be individually adjusted to the achievement level of the examinee. The computer moves the student selectively through a bank of items available at several levels of difficulty, bringing up an easier question after each wrong answer and a more difficult question after each correct answer. The benefits of this form of testing can be summed up in one word - efficiency. Such tests characteristically attain a specified level of measurement precision in about half the length of time a conventional test would require. They also ensure greater standardisation of administrative conditions than paper-and-pencil tests normally do, and, most importantly for diagnostic testing, provide almost immediate feedback to the student.

Since the early 1980s, considerable developments along these lines have taken place in the United States (McBride, 1985), in a variety of different assessment settings, including aptitude tests used in vocational counselling and selection, basic skill or competency testing, placement testing at the secondary and post-secondary levels, and diagnostic testing. Plans for the production of such tests in mathematics in the elementary operations of addition and subtraction have also been foreshadowed in Great Britain by the Director of the National Foundation for Educational Research in England and Wales (NFER), and research is currently in train at the Australian Council for Educational Research with the aim of preparing computer versions of existing ACER mathematics tests. This relatively recent development has the potential to be the most significant advance in diagnostic testing for many years, and may be the breakthrough necessary to ensure that effective error diagnosis in the basic skills becomes a reality in the classroom.

2 AIM OF EVALUATION

The aim of the present project was to determine whether a computer-based diagnostic program, modelled on Seville's diagnostic arithmetic tests, which have been well-known in New Zealand schools for many years, could assist a typical classroom teacher to locate problem areas and difficulties in the elementary arithmetic operations of addition, subtraction and multiplication for children at the Standard 3 and 4 level. The original research question was thus, 'Can the computer be used to diagnose problems in basic addition, subtraction and multiplication as well as a paper-and-pencil test does?' This included estimates of validity and reliability, to be obtained by comparing the accuracy of the diagnosis with that of an experienced remedial teacher and of a regular classroom teacher, together with some estimates of convenience and ease of use, time taken, and usability of the information supplied.

The argument was that if the program was successful in the diagnosis phase, and provided valid and useful information, with a minimum input of time, then more time would be available for the professional task of remediation. The teacher would be able to give the necessary individual attention, based on the results of the computer printout, focussing on the errors made, and not having to spend time working through a list of examples worked correctly.

No attempt was made in this study to test skills in the operation of division, as the division algorithm did not lend itself as easily to input from a computer keyboard, except for very simple examples. Several different ways of writing down intermediate answers were likely to be encountered, and the need to record 'carrying' figures on paper was likely to create additional difficulties. Given time, it would no doubt have been possible to prepare such a test, but it was felt that the three standard operations would give the computerized diagnostic process an adequate trial.

3 GENERAL PROCEDURE

The first phase of the experiment was administered in two classes in each of three Christchurch schools during Terms 2 and 3 of 1986, as follows:

Term 2: Somerfield Contributing School (Std 3, Std 4)
Redcliffs School (Std 3, Std 4)
Term 3: Elmwood Normal School (Std 3, Std 4)

Prior to their involvement in the project, each of the six class teachers had undergone a two-day computer familiarization workshop at the Christchurch Teachers College, during which they were introduced to the operation of the BBC Computer, and tried out the same introductory programs as their pupils would use, to allow them to get used to the keyboard. In fact, only two of the six teachers involved had had any prior computer experience, and so this familiarization phase was vital to the success of the experiment. The group was then introduced to the three subtests of the Seville Diagnostic Arithmetic tests, in both the paper-and-pencil and computer versions. They were also given some instruction on necessary computer 'house-keeping' matters, such as entering the class roll onto a computer file and recording results, as well as being given a briefing on the requirements of the evaluation, selecting the children for diagnosis, keeping logs of computer activities, and so forth.

Two computers were placed in each classroom for a period of 7 weeks, made up of a 5 weeks' familiarization phase, followed by a two week test phase. During the familiarisation phase, children were given the opportunity to experiment with a number of computer software programs, including interactive fiction *Flowers of Crystal*, *Dragon World*, *L,* (a fascinating mathematics adventure game requiring the correct answers to mathematical 'puzzlers' to progress), *Telebook* (a simple word processing program) and several public domain 'game' programs. With the occasional exception of *Telebook*, little attempt was made to use the computer activities as part of the normal classroom programme, but this was neither required or suggested.

For the test phase, 10 children were selected by the class teacher, on the basis of existing class records, Progressive Achievement Test results in

mathematics, and other information available, as those who would be most likely to require remedial work. They were not necessarily the 'bottom' 10 in the class in mathematics, although it was intended that they should fall within the 'bottom' half.

All the children in each class had the opportunity to attempt the computer diagnostic programs, in one or more of addition, subtraction and multiplication, but the 10 selected students were given the opportunity to do all three. In addition, they sat a paper-and-pencil version of the Seville Diagnostic Arithmetic tests. For 5 pupils, chosen at random, these were administered by their usual classroom teacher; for the other 5 pupils, they were administered by the itinerant remedial teacher. It was intended that the order of the 'treatments' should also be randomized, so that half the students in each school were exposed to the computer diagnostic program before the paper-and-pencil version, and vice versa, to keep practice effects to a minimum, but time constraints did not allow this to occur. However, teachers were careful not to teach any topics specifically reinforcing the material in the diagnostic tests, in the time interval between the computer diagnosis and the paper-and-pencil test. Neither the classroom teachers nor the visiting teacher had access to the computer scores of those children they tested subsequently, so that it was a true 'blind' trial. Any remedial action which followed from the study occurred *after* the second of the two assessments, on the basis of the results shown, and not *before*.

Pupils kept their own logs of programs attempted, how long they spent on them, and what they thought of them, and some obviously went to a great deal of trouble to produce elaborate and insightful records of their first 'computer experience'. The teachers also recorded their impressions of the experiment in a free-form log. A discussion of this log material is given in Section 5.

Hardware

The hardware requested and provided by the Department of Education was four BBC Microcomputers, complete with monochrome monitors and single disk drives, together with one SG 10 parallel printer with BBC interface. These were moved from school to school, as required. When the request was first made, it was assumed that the screens would be colour, but this proved not to be the case. Amber screens were provided, and although they do not seem to have affected the diagnostic programs (which were not designed for colour), the amber screens did limit the usefulness and ease of operation of some of the

introductory programs which were designed to make particular use of the colour facility.

Software

Software for the 5 introductory weeks within each school was provided by purchasing copies of *Flowers of Crystal, Dragon World, L* and *Telebook.* Assorted additional public domain programs were also provided to the schools. As so often occurs with this class of software, there were no instructions available. This was not seen as a great problem, as the purpose of this time was to make the pupils feel at home with the computer.

The software for the diagnostic period had been developed by the Christchurch Teachers College, and underwent further development as the project progressed. An account of the modifications which were made is included in the section on Formative Evaluation later in this report. In essence, the programs consisted of three modules, dealing with addition, subtraction and multiplication, constructed to parallel the various levels in the Seville Diagnostic Arithmetic Test.

The object was to present the pupils with 'user-friendly' testing programs, which adapted to their performance by moving them up or down through graded steps, according to the number of errors which they made.

Seville Diagnostic Arithmetic Tests

These diagnostic tests, devised by the headmaster of Manchester Street School, Feilding, and first published by the Australian Council for Educational Research in 1952, have been much-used in New Zealand in the diagnosis of errors in arithmetic. Each test provides for a hierarchy of levels of increasing complexity. An early version, with only two items per level, was used at Somerfield Contributing School; this was subsequently modified, after trialling at the remedial centre at Christchurch Teachers College, to one with four items per level, with the levels closely articulated with the levels of the computer version of the test. This version, which had been found to be a more satisfactory instrument for remedial work at the College, was used at Elmwood Normal School and Redcliffs Primary School.

The Sample Schools

Three schools in the Christchurch area were chosen for the evaluation, designed to show a range of teaching environments and reflect a variety of community catchment areas. The principals of each of them had some previous connection with the Teachers College, and the Elmwood Normal School in particular had close links through its requirement to provide demonstration lessons for College student teachers. But they were not atypical schools, in the sense that their teachers had had more exposure to computers, or were known to be particularly enthusiastic about having them in the school. They were certainly supportive of the idea, however.

Somerfield Contributing School: Somerfield is an older suburb, with varied housing styles. Moderately cheap, older housing has made it possible for young families and solo parents to move into the area. Children tend to be generally of 'average ability'.

The school has a staff of 15 teachers, which has remained relatively stable in recent years, with few changes. The Std 3 and Std 4 classes taking part in this computer study were housed in an open plan block of a design peculiar to Canterbury schools, popularly known to teachers as the 'Kentucky Fried' design because of the skylight peak over the centre of the building. Under this peak is an inner withdrawal or resource room. These rooms have been found to be particularly useful because of bench space, power points and proximity to each classroom.

Redcliffs School: Redcliffs is a seaside suburb some 10 km from the centre of Christchurch. It is a reasonably affluent community, with a large number of parents drawn from the professional classes, and little unemployment. Most parents own their homes, and many children have access to computers at home.

The school is a full primary school, with classes going up to Std 6 (Form 2). It has a staff of 12, teaching largely in relocatable classrooms. Because of the limited space in these rooms, it was necessary to place the computers on a bench in the back of each room about 1 metre apart.

Elmwood Normal School: Elmwood is set in the affluent northern Christchurch suburb of Merivale. This is a trendy address for the status-seeking, but at the same time is an established area of gracious houses and 'old' money. At the eastern and southern borders of the school zone, high-density apartments have led to a small transient school population. The majority of children

come from professional families, with high expectations of their children and the school. Parental involvement in school activities is strong and welcomed by staff. Children tend to be above average in ability, and are generally easily motivated. Behaviour problems are rare. The chief disadvantages are over-anxious parents and materially self-satisfied children.

The 1987 Replication

As the equipment was still available in 1987 on loan from the Department of Education to the Christchurch Teachers College, it was possible to carry out a replication of the study in 1987, involving the same teachers and schools, but different classes. The teachers were very enthusiastic to try the experiment again, but this time it was regarded as something of an 'extra' and was carried out more informally as a form of action research.

The three schools were not visited by staff from the Teachers College, other than to deliver and set up the equipment and make sure that all the hardware and software was present and functioning. The teachers themselves initiated the testing, for both the paper-and-pencil and computer versions, selecting their samples of 10 pupils per class and recording all the results without intervention from Teachers College staff. The remedial teacher from the Teachers College took no part in the exercise, which was treated entirely as if it were part of the normal classroom procedure in the schools. No teacher or pupil log books were kept, and the introduction and familiarisation process was left entirely to the teachers themselves.

The equipment was placed in Redcliffs School for Term 1 of 1987, in Elmwood Normal School for Term 2 and in Somerfield Contributing School for Term 3. At Redcliffs one of the two teachers who had participated in the 1986 experiment was allocated a Standard 2 class in 1987, and so it was decided to test this class and see whether the approach would work lower down the school. The remaining teacher administered the procedure to a Standard 3 class. Elmwood chose a Standard 3 and a Standard 4 class, as in 1986. Somerfield School made use of the computers, but not the test materials in either the paper-and pencil or computer versions. Staff there were very enthusiastic about having the computers for another term, but preferred to experiment with word processing instead. This school thus did not participate in the 1987 diagnostic arithmetic experiment.

4 ANALYSIS OF RESULTS

Effectiveness

The first question to be asked of the experiment relates to the accuracy of the computer diagnosis of errors, and attempts to answer the question of whether the computer provided a valid way of diagnosing errors in arithmetic. Results for Somerfield Contributing School are not included here, as it was the first school to take part in the experiment, and the version of the Seville Diagnostic Arithmetic tests used was not sufficiently well articulated with the levels employed in the computer program to allow detailed comparisons to be made, level by level. In a diagnostic test, comparisons of total scores have relatively little meaning; it is the comparisons at each level which are critical. There were also some difficulties with the hardware and the computer program itself which made the reliability of the results suspect. These are discussed in more detail in Section 5 on Formative Assessment.

After trialling at Somerfield Contributing School, both the paper-and-pencil version and the computer version of the Seville tests were modified to bring them more closely into line and improve the validity of the comparisons to be made. The final paper-and-pencil version is contained in Appendix A, along with a description of the operations involved in each of its progression levels. These generally follow Seville's hierarchy, with each level requiring more complex operations than the one before. There were 26 such levels in the Addition subtest, each consisting of 4 items of approximately equivalent difficulty, of which the first 2 were used in the test, with the remaining 2 being held in reserve for remedial purposes; the Subtraction subtest contained 20 such progression levels of 4 items, and the Multiplication subtest contained 21 levels.

The computer programs were designed to generate similar items at each level, using a random number generator to produce the digits within specified limits to ensure that the items would be comparable in difficulty to those in the corresponding levels in the paper-and-pencil tests. The Addition Module contained 25 levels, the Subtraction Module 16 levels, and the Multiplication Module 19 levels.

On beginning the Addition Module, each pupil was automatically entered at

Level 9, rather than at the beginning on Level 1. If both answers at this level were correct, the pupil skipped Level 10 and went directly to Level 11; if only one out of the first two answers was correct, the pupil was presented with a third example, to confirm the diagnosis. If this was correct, the pupil was also directed to Level 11, but the one incorrect example was flagged by the computer as a 'trip-up', for subsequent printout and remediation, if necessary. If the third answer was incorrect, or if the first two answers were both wrong (in which case a third example was not presented), the pupil was directed back a level to Level 8. If successful at this level (2 out of 2 or 2 out of 3 answers correct), the pupil was offered two new examples, randomly generated, at Level 9 again, and was able to proceed upwards once more. However, having failed two examples at Level 9 the first time, even if successful the second time, such a pupil would no longer move at the accelerated pace on odd levels only, but be presented with examples at every level from that point onwards.

Such a pupil, of course, would often fail at Level 9 again, and it was common for pupils to oscillate two or three times on two adjacent levels. In the early versions of the program, once a pupil had failed on six items, at any level, the diagnostic process was automatically terminated, and the session was ended. This gave a considerable amount of information about the difficulties they were facing at the point where the test became too hard, but relatively little information about their likely performance on other higher levels, which may in fact have been easier for them. A modification was therefore built into the program following the trialling at Somerfield School, which jumped the pupil ahead three levels from the last successful one, once they had made their six errors. A pupil succeeding on Level 8, but consistently failing on Level 9, thus had the opportunity to continue on Level 11, and similarly if they 'got stuck' at other points in the test.

It was commonly found that 'higher', supposedly more complex operations, were not necessarily more difficult; and that pupils having difficulty, say, with 'bridging' in addition, (renaming tens and ones, or 'carrying' figures) could proceed to problems with more digits, but not requiring 'bridging' and do them correctly. The modified program used in Elmwood and Redcliffs schools thus gave a greater range of useful diagnostic information through the incorporation of this 'jump' routine.

Throughout the computer session, every incorrect answer was automatically flagged, so that a printout of all incorrect answers, with the example which generated it, could be listed at the end of the experiment for the purposes of remediation.

The best pupils thus were able to progress through Levels 9, 11, 13, 15, 17, 19, 21, 23, 25, (9 in all) with no more than one error at any level; less successful pupils would be directed to the even-numbered levels from time to time, before resuming their upward climb. Some of them attempted as many as 13 different levels, sometimes more than once. Others felt they were not making progress, and after a period of inactivity, took advantage of a computer prompt which allowed them to 'give up' after having attempted no more than perhaps 6 or 7 levels. As this was a diagnostic test used in an experimental situation, no answers were given to the pupils as they progressed, and so success or failure on answered items would not be a factor in such a decision, although difficulty in getting any answer at all might have been.

On the Subtraction Module, pupils were started on Level 8, and proceeded rapidly upwards on the even-numbered levels 10, 12, 14, and 16 if they made no more than a single error at any level, but were re-directed to an odd-numbered level if they failed on two items at any stage, and then progressed more slowly. The minimum number of levels which might be encountered was thus only 5, compared with the mandatory 16 in the paper-and-pencil version.

The Multiplication Module started pupils on Level 6, from which they could proceed as far as Level 18 in similar fashion, moving up through the even-numbered levels, with the odd-numbered levels again being used as initial branching levels following failure. The minimum number of levels encountered was thus 7, compared with 19 in the paper-and pencil version.

Below are some examples of the diagnostic process, with the computer-generated output (not seen by the pupil of course), the error score calculated subsequently, and the parallel score given by the remedial teacher or class teacher on the paper-and pencil version in the far right column for comparison.

Pupil M.S.

Computer Output				Errors Detected Computer	Errors Detected Tester
0	LEVEL 6	Correct		0	0
1	LEVEL 8	Correct		0	0
2	LEVEL 10	Correct		0	2
3	LEVEL 12	Correct		0	1
4	LEVEL 14	Correct		0	0
5	LEVEL 16	: 237x4=928		1	2
6	LEVEL 18	Correct		0	0
7	LEVEL END	6 min 33 sec			
			TOTAL	1	5

The first is a case of a reasonably capable student, Pupil M.S., suffering a single 'trip-up' in the computer version of the multiplication test, on Level 16, and so passing straight through quite quickly, but making more errors in the same levels of the paper-and-pencil version.

This next pupil (G.M.) failed level 16 the first time through, and was re-directed to Level 15, on which she succeeded with 2 correct responses, and then was successful at Level 16 on the second try. The errors on her first try at Level 16 are noted for diagnostic purposes, but not for the purposes of compiling the total error score, which remains at 0. She made no errors at all on the paper-and-pencil version, also scoring 0.

<u>Pupil G.M.</u>

Computer Output			Errors Detected: Computer	Errors Detected: Tester
0	LEVEL 8	Correct	0	0
1	LEVEL 10	Correct	0	0
2	LEVEL 12	Correct	0	0
3	LEVEL 14	Correct	0	0
4	LEVEL 16	: 876-7=860: 147-8=149	-	-
5	LEVEL 15	Correct	0	0
6	LEVEL 16	Correct	0	0
7	LEVEL END	6 min 51 sec		
		TOTAL	0	0

The next more complex example (Pupil P.L.) contains a wealth of diagnostic information about the difficulties which the pupil is having in handling 'bridging' of three-digit numbers, but it is not the object of the present evaluation to engage in error diagnosis at this point. It is interesting to note, however, that some learning is taking place, and that eventually the pupil succeeds on Level 22 and Level 23 (although not Level 24). For the purposes of the analysis to follow, the smallest number of errors on any level is recorded, and the pupil thus registers 4 errors, 'trip-ups' on Levels 11 and 21, and a failure on Level 24. Clearly some difficulties are being experienced at Level 22 and 23, though, and the error patterns can give helpful information on this.

The results from the paper-and-pencil version of the test gave a fail at Level 23 and a pass at Level 24, but a generally similar diagnosis overall. This pupil failed to follow instructions, and terminated the program in an abnormal way, probably by using the BREAK key; this is likely to be the reason why no time was recorded.

Pupil P.L.

Computer Output			Errors Detected Computer	Tester
0	LEVEL 9	Correct	0	0
1	LEVEL 11	: 59+9=86	1	0
2	LEVEL 13	Correct	0	0
3	LEVEL 15	Correct	0	0
4	LEVEL 17	Correct	0	0
5	LEVEL 19	Correct	0	0
6	LEVEL 21	: 134+445=578	1	0
7	LEVEL 23	: 172+269=534: 186+338=411	-	-
8	LEVEL 22	: 518+125=0 : 119+112=411	-	-
9	LEVEL 21	: 172+525=787:	-	-
10	LEVEL 22	: 449+136=57 : 218+167=37	-	-
11	LEVEL 21	: 115+274=388	-	-
12	LEVEL 22	Correct	0	0
13	LEVEL 23	: 127+786=91	-	-
14	LEVEL 24	:689+448=1127: 599+414=913	2	0
15	LEVEL 23	Correct	0	2
(PROGRAM TERMINATED - NO TIME RECORDED)				
		TOTAL	4	2

Validity: Total Number of Errors

Results in Table 1 show the mean number of errors per student, as detected by the computer program, in comparison with the number of errors found by the visiting remedial teacher *on the same levels*, for each of the subtests in each of four classes in 1986. It should be emphasised that the comparison only applies to the common levels attempted, which could be different for each pupil, depending on the particular route which each one took through the various progression levels in the computer version of the tests.

It is apparent from the means shown that there is a slight tendency for more errors to occur in subtraction and multiplication, rather than in addition examples. But there is only one statistically significant difference between the computer results and the remedial teacher results, although the small sample sizes mean that a difference would need to be quite large to exceed even the $p < .05$ level, the minimum normally regarded as acceptable for statistical significance in such studies. The significance test used was a *t*-test for correlated samples, since the same randomly-chosen groups of 5 pupils from each class were administered both modes of the test.

Table 1

Comparison between mean number of errors detected by computer and by remedial teacher on common levels of modified Seville Diagnostic Arithmetic Tests: 1986

Class	Computer	Remedial Teacher	Sig. Diffs.
Elmwood Std 3: (N=5)			
Addition	1.6	2.4	
Subtraction	2.8	1.6	
Multiplication	6.4	5.6	
Elmwood Std 4: (N=5)			
Addition	1.6	0.8	
Subtraction	1.8	1.8	
Multiplication	3.6	0.8	+
Redcliffs Std 3: (N=5)			
Addition	3.6	2.0	
Subtraction	3.8	4.4	
Multiplication	3.6	5.4	
Redcliffs Std 4: (N=5*)			
Addition	2.8	1.6	
Subtraction	2.0	2.6	
Multiplication*	1.5	4.8	

* Only 4 students were tested by the remedial teacher in multiplication.
\+ $p < .05$

Table 2 presents similar information on a comparison between the scores on common levels from the computer and the class teacher for 1986. The figures in the table are again the mean number of errors on the common levels of each subtest, as detected by the computer and the class teacher. Once again, there is only one significant discrepancy, and the only obvious trend is for the computer diagnostic program to detect more errors in the addition subtest, in each school.

In sum, it could be said that, when assessed by the fairly gross measure of the total number of errors detected, the computer program does as well as the paper-and-pencil version of the modified Seville Diagnostic Arithmetic Tests, administered either by the remedial reading teacher or a classroom teacher. A separate comparison of results between the visiting remedial teacher and the class teacher gave no significa t differences, and these would not be expected, as the test administration was standardized and gave little room for variations in procedure.

Table 2

Comparison between mean number of errors detected by computer and by class teacher on common levels of modified Seville Diagnostic Arithmetic Tests: 1986

Class	Computer	Class Teacher	Sig. Diffs.
Elmwood Std 3: (N=5)			
Addition	1.8	1.0	
Subtraction	1.8	2.0	
Multiplication	2.8	2.6	
Elmwood Std 4: (N=5*)			
Addition*	2.0	1.0	
Subtraction	1.8	2.4	
Multiplication	2.0	3.8	
Redcliffs Std 3: (N=5*)			
Addition	3.4	1.8	+
Subtraction	3.0	2.8	
Multiplication*	2.8	1.8	
Redcliffs Std 4: (N=5)			
Addition	2.2	1.2	
Subtraction	1.8	2.0	
Multiplication*	4.0	1.2	

* Only 4 students were tested by the class teacher in each case.
\+ $p < .05$

Results from the 1987 Replication

Prior to the 1987 year, the computer programme had undergone further 'fine-tuning' by the Teachers College staff, to make it even more sensitive, and additional modifications were made to the way in which pupils progressed through the various levels. Some of the algorithms were altered, so that pupils encountering failure at a particular level were moved ahead several levels to examples of a different kind or in a different format which they may have been able to do. For example, pupils having difficulty in handling 'bridging' operations in two digit addition were moved on to questions involving three digits, but no 'bridging'; pupils striking trouble in renaming and bridging in multiplication of tens by ones could be moved on to three column multiplication (hundreds by ones) without renaming. The general structure of the computer version was, however, maintained, with pupils dropping back a level if they failed at a higher level, until the session was terminated or they decided to give up. In the 1987 version of the Addition

Module, pupils started on Level 4 rather than Level 9 as in 1986, and progression was not simply through alternate levels, but the size of the 'jumps' depended on the type of example being presented. In the Subtraction Module all pupils started on Level 2, and were asked to attempt every level for a while, before being allowed to jump. In the Multiplication Module they began on Level 4. It was hoped that this new version of the program would give more comprehensive information by presenting just sufficient examples to pupils at the point where they were finding difficulty to provide for sound diagnosis but not generate 'overkill', and still allow as many as possible to continue to the end, where more advanced, but not necessarily more difficult, examples were located.

Table 3

Comparison between mean number of errors detected by computer and by class teacher on common levels of modified Seville Diagnostic Arithmetic Tests: 1987

Class	Computer	Class Teacher	Sig. Diffs.
Elmwood Std 3: (N=9-10)			
Addition	3.7	0.8	+
Subtraction	7.8	4.1	+
Multiplication	7.6	4.7	+
Elmwood Std 4: (N=8-10)			
Addition	4.5	1.5	+
Subtraction	9.4	4.4	+
Multiplication	8.3	6.4	
Redcliffs Std 2: (N=7-8)			
Addition	3.9	0.8	++
Subtraction	7.1	5.4	
Multiplication	5.9	5.4	
Redcliffs Std 3: (N=6-9)			
Addition	5.3	2.7	
Subtraction	8.9	5.6	
Multiplication*	8.1	7.4	

\+ $p<.05$
++ $p<.01$

The results in Table 3 from the 1987 trials show some significant differences, with the computer program regularly detecting more errors than the pencil-and-paper version administered by the classroom teacher. This may be because the 'fine-tuning' which had taken place in the program modules made

them more sensitive to the types of errors which pupils were making, and allowed them to jump more flexibly to new examples rather than 'drop out' when they made a series of mistakes. Pupils were also started nearer the beginning in the 1987 versions of the programs, and would thus attempt more examples and have more opportunity to make errors. It is also possible, of course, that in the relatively unsupervised environment of the 1987 trials, pupils were not made sufficiently familiar with the computers before they tackled the arithmetic programs, and were making errors in data entry unrelated to their knowledge of arithmetic. Some evidence to be presented later in Section 5 suggests that a few large discrepancies may have been caused by a failure to follow the on-screen instructions properly in the entering of answers. These 'outliers' will have inflated the mean number of errors.

Nevertheless, the fact that the computer program is detecting more errors, and every one of these errors is documented and available to the teacher for remedial purposes, suggests that it is likely to be of considerable assistance in the classroom.

Validity: Correlations

A check on the validity of the computer assessment process was also made by correlating the number of errors detected by the computer for each pupil with the number of errors detected for the same pupil by either the remedial teacher or the class teacher, again over only the common levels on the tests. Summary results are given in Table 4. The correlations are not high, particularly in 1986, and a glance at the corresponding scatter plots for the

Table 4

Correlation coefficients between number of errors detected by computer and by testers on common levels of modified Seville Diagnostic Arithmetic Tests

Subtest	Remedial Teacher 1986	Class Teachers 1986	Class Teachers 1987
Computer:			
Addition	0.47	0.22	0.45
Subtraction	0.28	0.45	0.44
Multiplication	0.19	0.06	0.57

NOTE: The sample numbers upon which these correlations are based range from 18 to 20 for the 1986 study, and from 33 to 35 for the 1987 replication.

1986 samples in Appendix B suggests that the reason is probably related to the small number of errors generally being recorded by the computer program, and to their relatively broad scatter.

The larger number of errors recorded by the computer programs in 1987 may perhaps have led to a more obvious relationship, reflected in the size of the 1987 correlations. These figures are based on larger sample sizes, of course, since the remedial teacher was not involved and class teachers were responsible for all the testing in that year.

Perhaps a better method of checking the accuracy of the process is to record for each pupil in how many separate levels the number of errors found by the computer agreed exactly with the number found by the remedial teacher or classroom teacher, in how many levels the computer detected more errors, and in how many it detected fewer. Perfect agreement would be likely to occur for some pupils, but not all. However, if the level of agreement was relatively high overall, and there was no tendency for the computer to over- or under-estimate, the process could be regarded as valid. Detailed results for each pupil for each test in the 1986 study are to be found in Appendix C, and a summary is given for both years in Table 5. The numbers in the table are the *percentages* of levels falling into each category, averaged across all pupils.

The results display a good correspondence in the addition and subtraction subtests in 1986, with the mean percentage of levels showing an exact correspondence between the computer and paper-and-pencil versions of the tests being quite high. The correspondence in the multiplication test is somewhat lower, but even here there is agreement in nearly three-fifths of all the common levels, averaged across pupils. These results support the lower correlation coefficients found for multiplication in Table 4.

In the subtraction and multiplication subtests for 1986 there is no overall tendency for the computer to either under- or over-estimate the number of levels on which errors have occurred; in the addition subtest the computer is locating more levels containing errors than the paper-and-pencil test. This confirms the diagnosis on the basis of total number of errors, as reported in Table 2.

In 1987 the computer is consistently finding errors on more levels than are the classroom teachers, in all subtests, but the 'exact match' column remains reasonably high, particularly when the results of Redcliffs Standard 2 pupils are omitted. These children found both the paper-and-pencil test and the computer version too difficult for them. They were put off quickly, gave up when they could not understand what to do, and generally only completed a

few of the levels. Their scores are thus somewhat unreliable, and the results in the last section of Table 5 leaving them out show a better match. Even so, the computer version is clearly flagging more levels for attention.

Table 5

Mean percentage of levels in the Seville Diagnostic Arithmetic tests in which the number of errors detected by computer matched the number detected by testers

Subtest	Computer more %	Exact match %	Computer fewer %
1986:			
Addition	15.9	76.8	7.3
Subtraction	11.5	75.3	13.1
Multiplication	22.1	58.6	19.3
1987:			
Addition	28.5	67.3	4.1
Subtraction	28.4	64.0	7.6
Multiplication	28.2	55.3	16.5
1987 (Excl. Redcliffs Std 2):			
Addition	20.7	74.1	5.2
Subtraction	27.6	65.4	7.2
Multiplication	26.4	58.2	15.4

Efficiency

Two simple measures were used to assess the efficiency of the computer program against the paper-and-pencil version of each subtest. The first is the total time taken to completion; the second is the mean time per level attempted. Not all teachers kept consistent records of the total time for each pupil to complete the paper-and-pencil version of the test in the 1986 trials, and so only results from the remedial teacher are presented, recorded to the nearest minute. The computer program logged the total elapsed time to the nearest second, for every pupil who obeyed instructions and allowed the program to terminate normally. The mean time per level was calculated on the actual number of levels completed for each subtest. For the paper-and-pencil versions, this was assumed to be the total number of levels in the test, as very few pupils failed to complete all items. For the computer version, it was the total number of levels attempted, counting separately all repeated levels, and not simply the final highest-scoring level which contributed to

the total error score. The results for both measures are given in Table 6. No time data were collected in 1987.

In general, the computer performed its diagnosis in a shorter time than it took the remedial teacher to administer the paper-and-pencil test, except in addition, although the data for this subtest were somewhat limited. The reason for this has already been noted; some pupils became bored or tired of trying, and terminated their sessions by using the BREAK key, rather than by allowing the program to proceed to the end and finish normally.

Table 6

Comparison between mean time taken by computer and by remedial teacher in administration of modified Seville Diagnostic Arithmetic Test (seconds): 1986

Class	TOTAL TIME Computer	Teacher	Sig. Diffs.	TIME/LEVEL Computer	Teacher	Sig. Diffs.
Elmwood Std 3: (N=5)						
Addition	-	660		-	26	
Subtraction	826	756		74	38	+
Multiplication	1064	1242		90	59	+
Elmwood Std 4: (N=5)						
Addition	-	576		-	23	
Subtraction	616	804		94	40	
Multiplication	-	1068		-	51	
Redcliffs Std 3: (N=5)						
Addition	930	612		71	24	
Subtraction	436	996	+	71	50	
Multiplication	434	1308	+	71	62	
Redcliffs Std 4: (N=5)						
Addition	885	528		74	21	++
Subtraction	741	840		86	42	
Multiplication	676	1050		79	50	

\- Indicates that time data were available for fewer than 4 out of 5 students.
\+ $p<.05$
++ $p<.01$

When considered on a time-per-level basis, the computer took rather longer than the paper-and-pencil test. But as one point in the exercise was deliberately to *reduce* the number of levels presented to a pupil who was progressing without difficulty, and *increase* the number of items presented to pupils finding difficulty, at the point where they first began to experience

failure, this latter measure has less relevance. The total time to complete the test has more significance as a measure of efficiency.

Furthermore, the diagnostic process on the computer could proceed without the constant supervision or intervention of the teacher. This fact, along with the availability of a detailed printout of every problem which a child got wrong, for subsequent error diagnosis, are undoubtedly the chief advantages of the computer program, features which made it so appealing to the teachers.

5 FORMATIVE ASSESSMENT

In any educational innovation, the particular context in which the experiment occurs is bound to have an impact on the outcome, and so upon the conclusions which can legitimately be drawn. Educational research does not occur in a vacuum, and the particular community environments, the expectations and competencies of principals, teachers and pupils, as well as the performance of the computer hardware and software will have an important effect. This section considers some of these environmental matters, describing briefly the particular settings in which any problems occurred, the remedies which were attempted, and the general impressions of both teachers and pupils about the experiment, drawn from information contained in their computer logs.

General Hardware Problems

Some problems were experienced with the hardware which was supplied for this project. A screen proved faulty, and had to be returned. It was repaired, but the display was rendered less bright than originally, and could not be improved. This caused little problem in a semi-shaded situation, but was to become a nuisance later in classrooms.

Two disk drives had to be returned at different times. One was not operating correctly when it was received, and it was returned and replaced. The other caused greater problems. It appeared to be operating correctly, and it was not until it was being used at Redcliffs School that it was found not to be writing onto the disk. The lack of some results from the school which had used it previously, Somerfield, was thought to have occurred for other reasons. When it was discovered that a number of pupils had apparently not done their tests (to the surprise of their teacher!), a short program was written to enable the teacher to check the names of those who had done the test. This confirmed the source of the problem, and the disk drive was withdrawn.

There was also a problem with one keyboard. Although the computer appeared to be operating correctly, a variety of unusual sounds came forth when different keys were pressed. This keyboard had to be returned to Auckland for repair. Fortunately for the study, a sufficient number of

similar computers owned by the Christchurch Teachers College itself were available for temporary loan, and the situation did not occur in which fewer than three computers were available in the trial schools; four were usable for most of the time.

In spite of these defects, the BBC hardware was generally deemed satisfactory, but an obvious weakness did become apparent when computers had to be moved from room to room and school to school, however much care was taken. There were problems with plugs, especially those linking the disk drive to the computer. They did not fit tightly to begin with, came out easily, and could be difficult to replace. Not all the problems were directly the fault of the hardware. When a disk drive is accidentally knocked off a table and hangs by the leads, trouble can be expected, but it was rather frustrating to have to dismantle the whole drive in order to put a plug back in!

Software Modifications

The central software for the project has already been described, and because it was written in BASIC at the Christchurch Teachers College, by one of the authors of this report, it was possible to modify it during the conduct of the study. This 'fine tuning' was an intended outcome of the investigation, and continued over the two years of the evaluation.

Two design features created problems at the outset, and needed attention. One was concerned with the length of time a pupil should be left sitting at the computer without recording a response. This was handled by including three prompts which appeared at appropriate intervals. They were: 'Enter a number or press RETURN', 'Is this too hard?', and 'Do you want to stop?'. The real question here was what was the appropriate time interval between each prompt. This was determined empirically, after observing a number of pupils who were finding difficulties in knowing what to do. The program delays were adjusted accordingly.

The other more serious problem related to whether answers should be entered with the digits running from left to right, or from right to left. It was finally decided that when a problem was presented in horizontal form, e.g.

$$5 + 14 =$$

that answers should be entered with the digits running from left to right (i.e. 1 then 9 in this example) as in this case one would expect a pupil to

verbalize the answer as '5 plus 14 is 19' and enter 19.

On the other hand, when a problem is presented in vertical form, e.g.

```
  27
+ 16
  --
```

it is likely that the digits would be entered from right to left, as one would expect a pupil to begin '7 plus 6 is 13' and enter the 3 first, followed by the 4 after further calculations had been done. The problem was partly resolved by putting in a check. When the answer was entered, the computer asked, 'Is that what you really want?', and the pupil had an opportunity to correct an answer before going on to the next problem.

There is some evidence, however, that this prompt was not completely successful in avoiding reversals, particularly in the 1987 replication. Standard 2 Pupil C.M. has clearly not understood the order in which digits should be entered, noticeably in LEVEL 8, although she knows the answers to the addition sums, as shown by her scores on the paper-and-pencil version.

Pupil C.M.

Computer Output	Errors Detected: Computer	Errors Detected: Tester
0 LEVEL 4 : 4+7=21	1	0
1 LEVEL 5 : 41+1=15	1	0
2 LEVEL 8 : 92+5=79 :83+2=58	2	0
(PROGRAM TERMINATED - NO TIME RECORDED)		
TOTAL	4	0

Another more dramatic illustration is Pupil G.N., who has not really been able to come to grips with the order-of-digits problem at all well. For this pupil, reversals occur on LEVELS 8, 10, 12, and 13, in some cases along with other minor errors, and it is probable that the perfect score on the paper-and-pencil version of the test is the more accurate estimate of his abilities in addition. The discrepancy between the number of errors detected by computer and the number detected in the paper-and-pencil test is quite large. Just one or two discrepancies of this order can distort the means on small samples considerably, and this seems to have occurred more frequently in the 1987 replication of the experiment.

Pupil G.N.

Computer Output			Errors Detected Computer	Tester
0	LEVEL 5	Correct	0	0
1	LEVEL 8	: 82+3=58 :93+4=79	2	0
2	LEVEL 7	: 56+2=64	1	0
3	LEVEL 10	: 26+6=23 :29+3=33	2	0
4	LEVEL 9	: 28+4=33	1	0
5	LEVEL 12	: 57+3=6: 85+6=19	2	0
6	LEVEL 11	Correct	0	0
7	LEVEL 13	: 2+8+3=31	1	0
8	LEVEL 14	: 2+7+5+13: 8+9+6=22	-	-
9	LEVEL 13	Correct	0	0
10	LEVEL 14	Correct	0	0
11	LEVEL 16	Correct	0	0
12	LEVEL 17	: 33+53=8: 51+44=60	2	0
13	LEVEL 16	Correct	-	-
14	LEVEL 17	: 42+55=67: 21+38=49	-	-
15	LEVEL 21	Correct	0	0
18	LEVEL END	21 min 36 sec		
		TOTAL	11	0

Another illustration of a problem related to the computer administration of the test is shown by pupil J.D.

Pupil J.D.

Computer Output			Errors Detected Computer	Tester
0	LEVEL 2	: 11-8=4	1	0
1	LEVEL 3	Correct	0	0
2	LEVEL 4	: 98-12=0 :36-13=27	-	-
3	LEVEL 3	Correct	0	0
4	LEVEL 4	: 98-86=0	1	0
5	LEVEL 5	Correct	0	0
6	LEVEL 6	Correct	0	0
7	LEVEL 7	Correct	0	0
8	LEVEL 8	Correct	0	0
9	LEVEL 9	Correct	0	0
10	LEVEL 11	: 78-69=68: 92-86=0	2	0
11	LEVEL 10	: 75-18=0: 66-48=0	2	0
12	LEVEL 9	: 21-7=15	-	-
13	LEVEL 11	: 56-47=0: 31-28=12	-	-
14	LEVEL 12	: 40-13=37: 60-43=0	2	0
15	LEVEL 13	Correct	0	0
16	LEVEL 15	Correct	0	0
17	LEVEL 16	Correct	0	0
18	LEVEL 18	: 822-469=0: 913-659=0	2	2
19	LEVEL 17	: 678-49=0: 682-55=637	2	0
20	LEVEL 16	: 462-8=0	-	-
21	LEVEL END	27 min 12 sec		
		TOTAL	12	2

The large discrepancy between the two versions of the test is caused by the number of zero answers, probably generated by simply pressing the RETURN key without entering a number. This pupil is having difficulty with 'bridging' in subtraction, and is not handling it well. There are glimmers of understanding of the process, as in the first answer in LEVEL 12 and the last answer in LEVEL 17, but the computer presentation is clearly causing problems which the paper-and-pencil version is not. The difficulty in writing down 'carrying' figures while working on the screen may be the trouble, leading to guessing and incomplete answers.

Relatively few pupils had major difficulties of this nature however; generally they appeared to adjust to the novel form of administration without too much trouble. Quite a number, particularly the less confident ones, wrote their answers down on paper first, before keying them in. But these few aberrant results suggest that a little guidance from the classroom teacher at the outset would be desirable to ensure that the data entry procedures and conventions are fully understood by all children.

Somerfield Contributing School

The Std 3 class at Somerfield contained 33 children of mixed ability, with approximately equal numbers of boys and girls. The computers were placed alongside each other in a bay in the classroom, where it was reported that they caused very little disturbance. It was convenient to house them there so that children could be given assistance in the initial stages of their computer activities, when the rest of the class was busy on other work.

The class had had plenty of successful group experience before the introduction of the computers. A deliberate effort was made by the teacher of this class to pair children who might not normally have chosen to work together, but no problems were reported. Indeed it appeared to result in an improvement in relationships, and certainly improved group interaction and discussion. A small group of children were trained to handle the equipment, and they were on call if any of the other members of the class had difficulties. This resulted in minimum interference to the class programme. The only problem reported by the class teacher was the need to explain other work to children who had missed it while they were out of the classroom using the computer.

The 36 pupils in Std 4 were used to working individually and in groups, and the classroom programme needed no major changes to accommodate the study. Children were given timetabled days and times throughout the day, plus extra

times they could book, before or after school and at lunch times. The carpeted classroom and acoustic ceiling tiles were definite advantages, allowing the children to move freely from tables or floor to the computers without disturbing other pupils unduly. Children from a neighbouring composite Std 3/Std 4 class were also introduced to the computers, paired with experienced children initially. Only two or three of the 36 children had computers at home; a few more had access to computers in offices or in the homes of their friends.

Although the experiment went reasonably well according to plan in this, the first of the schools to try the new equipment, there were some difficulties which should be noted. As previously mentioned, there were a few problems with hardware, and these had adverse effects upon the results that were obtained. The two teachers involved were also the more senior of the six in the three experimental schools, and one of them, in particular, found that responsibilities in the school reduced the time that he had to devote to the project. On one occasion when the field workers visited his school they found him trying to cope with a shortage of six members of staff absent for the morning! This did not appear in any way to lessen his interest in the project, but it may help to explain why there seemed to be less enthusiasm and personal involvement there than in the other two schools.

Even though some problems were experienced at Somerfield, it was felt desirable to leave all arrangements to the teachers involved, as it was the intention to allow the scheme to operate in a 'normal' school environment, with all its pressures and constraints. The researchers made sure that the teachers knew what was required, and then did not intrude, but left them to cope with the various eventualities which might (and did) arise.

The results from Somerfield School suggested two things which are likely to affect the validity of the study. First, it seems that some pupils did not take the testing very seriously, and secondly, the time delay between computer prompts turned out to be too short. These problems were probably related to each other. To help overcome this in other schools, teachers were asked to explain carefully to each class that the computer was keeping a record of their results, while they worked away. The Somerfield School pupils may not have known this. During the introductory period when they were playing computer 'games' no records were kept, and it didn't matter if a mistake was made, and keys were pressed at random. Perhaps this influenced the way in which results were entered during the final week, when it *did* matter. Another influence may have been the wording of the first prompt. The instruction 'Enter a number' could have been interpreted to mean 'Enter any number'. This

was subsequently changed to 'Enter your answer'. The delay between prompts was also lengthened, to allow more time for the pupils to respond, without being reminded.

Redcliffs School

Classroom organization at Redcliffs School was stated by the teachers concerned to be a combination of individual, group and whole class work, but with a strong emphasis on individual work. Children were encouraged to work quietly, independently and to keep 'on task', in fairly formal seating arrangements. Core subjects were scheduled in the morning, and cultural activities in the afternoon. The children worked in pairs, timetabled into half-hour sessions throughout the day. The first class to participate in the study was a composite Std 3/Std 4 containing 28 pupils, of generally high ability, although containing 8 'below average' Std 4 pupils. One child was Indian, the rest of European origin. The other class was a Std 3 containing 34 children with a wide range of abilities, and some children with special needs. It contained one Japanese pupil, and one of Indian ethnic origin. Apart from the hardware problems already noted, everything went according to plan. The teachers were enthusiastic and used their time to work with their pupils in interesting ways. Perhaps the need to be careful was stressed too much, or perhaps too much was made of the fact that pencil and paper could be used to help work out the answers to problems, before typing them on the keyboard. Whatever the reason, some pupils took a long time to complete some of the tests, and this will have reduced the apparent 'efficiency' of the method, in comparison with the more formal administration under the control of the class teacher or the visiting itinerant teacher.

Once again, lessons were learnt which allowed further 'fine-tuning' of the tests. A pupil having difficulty with:

```
  30
-  8
  --
```

(a problem involving subtracting from zero) could go on in the paper-and-pencil version of the test and get

```
  58
- 35
  --
```

correct, even though this came from a higher level in the Seville test. But in its earlier version, the computer program was recognizing the first error,

and cutting them off at this point, without giving them the opportunity to jump ahead and attempt other items at supposedly higher levels which they may nevertheless have been able to do correctly. Adjustments were made to the software, by building larger 'jumps' into the program to move pupils from a level at which problems were being experienced to a higher level which tested different skills. The results of the analysis add further light on this point, and suggest that there is not a strict hierarchy of difficulty in the items at the various levels, although the processes appear to become more complex as the number of digits being handled increases. Further fine-tuning of the program took place before the 1987 replication, as has already been noted in Section 4, to attempt to optimize the amount of diagnostic information obtained.

Elmwood Normal School

The Std 3 class at Elmwood consisted of 18 boys and 13 girls; the Std 4 class of 15 boys and 18 girls. In the Std 3 room the computer centre was in a partitioned area in the back of the room; in the Std 4 room it was placed in the 'maths corner' to one side of the front blackboard wall. A common problem of computer noise during quiet class periods led to the removal of computers to an adjacent small classroom used by staff and children as a withdrawal room.

Class organization was normally based on curriculum studies in the morning, and a topic-related cultural activity in the afternoon. As Elmwood is a Normal School, there is a close association with Christchurch Teachers College, and the children were used to new faces, fresh ideas and a variety of teaching techniques. Whole class and group teaching methods were commonly used.

The children were reported as being enthusiastic, and interest in computers was not restricted to the brighter children. Participation was widespread, with computer use being regarded more as a function of experience. The large majority of pupils had a home computer or regular access to one. The Std 3 girls appeared to use the computer more often for process writing, in the introductory phases, while the boys tended to prefer the games disks. The Std 4 children generally did not use the printer or *Telebook*; the more able liked playing the more complex games, such as *Flowers of Crystal*, but no sex differentiation was noticed by the teachers at this level. In their view, children in both classes benefited from paired-learning situations.

Once again there were minor hardware problems, but the experiment was

carried through successfully. The teachers were enthusiastic and more knowledgeable about the operation of computers than at the other two schools. When the testing was about to begin, at least one of the classes was told something like 'If you have difficulties or the program goes on too long, press ESCAPE or BREAK'. Some pupils did this, and so 'dropped out' of the testing program too soon, causing the loss of some results (in particular the 'time taken' measure calculated automatically by the computer) and causing other odd things to be written onto the disk. This was solved later by disabling the ESCAPE key. The modified 'jump' instructions in the program appear to have reduced the problems which occurred when a pupil 'oscillated' between two adjacent levels, and couldn't get beyond them. The further development of the programs which took place before they were used again for the 1987 phases of the experiment was designed to improve their efficiency in this regard.

Children's Computer Diaries

All the children in the six classes participating in the experiment in 1986 were asked to keep a diary of what they did during the familiarisation phase of the experiment. A few carried on and wrote about the testing phase as well. Teachers generally gave some guidance about setting up a suitable format, but the children were left free to shape their diaries according to their own preferences. Every class produced something different, and some very elaborate and attractive records were submitted, although it was reported that they needed prompting to keep them up-to-date. The following suggestions about keeping their diaries were given to the children at the beginning:

Here are some things you might like to write about:

a. Did you enjoy using the computer today? Were you able to do what you wanted to?

b. Did you have any problems? What went wrong? Did you work out what you had to do in the end?

c. Did anyone work with you on the computer? What help were you given? Did you help anyone else? What sort of help did you give?

d. What can you do on the computer that you cannot do any other way? Do you prefer using the computer, compared to other ways of doing things?

e. Keep a brief account of the different things you use the computer for, and see if you improve your skill

from week to week or month to month. Can you tell if you are getting better?

There was absolutely no question about the fact that the children enjoyed what was, for a good number of them, their first computer experience. Adjectives like 'fun', 'exciting', 'neat', and even 'excellent', mighty', 'terrific', 'superb' are peppered throughout virtually every diary. This was particularly so for the games, rather less so for the diagnostic arithmetic modules. However, some pupils tempered their enthusiasm with more thoughtful, qualified commendation; some were frustrated at not making progress on the games; a few found them boring after a while and wished for more variety; and a few were critical of the various software infelicities and hardware faults already noted. A representative sample of evaluative comments follows:

The study was OK. I didn't think it was great but it wasn't bad.
(Philip)

I enjoyed having the computer. I hope we have the computer another time.
(Meredith)

It's good how you are learning while you're playing games. [Dragon World]
(Joanne)

It was quite hard, but very exciting, and also fun.
(Gayle)

The computers are very very excellent. I am saving up for one myself because I liked the ones at school so much.
(Karl)

I think that the experiment might do some good. A bit boring after a while. [not enough games]
(Emma)

Annoying when it said, 'IS THIS TOO HARD?' when you were working it out [The time delay for this prompt was modified subsequently]
(Teall)

The computer is rude, ignorant and needs to go to school. Words like 'to go' it does not understand. [This comment may be a little 'tongue in cheek', because the pupil making it rated the game "Reversi' superb, excellent, terrific!]
(no name)

I felt frustrated. [couldn't solve games]
(Daniel)

Whenever we move the computer something goes wrong. That means that some people miss out, and I'm usually one of them.
(Beth)

I enjoyed having the computers in our classroom. The noise was a bit annoying but we soon got used to that.
(Naomi)

Generally the arithmetic tests passed muster, and the children found them not too difficult to handle, although the absence of paper-and-pencil for intermediate working ('carrying' figures) proved a problem for some children.

At first the computer testing rushed me, but I got used to the pace. And it is more enjoyable then normal maths.
(Roland)

I think that they [arithmetic tests] were very easy. I wish it was a bit longer then it would really get your brain working.
(no name)

Doing maths on a computer is far less tiring but it makes me slightly nervous.
(no name)

I think computers are great fun, but I think it is easier to use paper than the computers because you can't carry your numbers.
(Jeffrey)

The multiplication test was harder than the others. I think I like the games better than the tests.
(Beth)

The maths tests we did on the computers were a lot easier than ones on paper because not everyone in the class is doing it and you have all the time you like.
(Sarah)

It was very obvious that many children saw the experiment as a valuable learning experience, both in mastering a new skill with the keyboard, and also in co-operating with other children in new ways. The pairing of children to work together on the computer brought about an appreciation of what it was to be ultimately 'in charge', pressing the keys and controlling the whole operation, and what it was to co-operate as an assistant giving advice to the one who had 'hands on'.

On Monday 16 June I had another go on the computer. The partners I had this time let me touch the keyboard more.
(Matthew)

The bell rang so I had to stop. Everyone was crowded around me, and told me what to do but I did not listen.
(Melissa)

On Wednesday the 30th May I had a go on the computer with Andrew ... He was a bit bossy but I managed to cope with him.
(Holly)

> I knew what I was doing this time.... I could tell I was getting better because I knew where the keys were.
> (Sally)

> The people next door on the other computer had a bit of trouble so we helped them. We had no trouble at all.
> (Trudi)

> I think it is good to work in pairs because in some games it's hard to make up your mind and you need someone to help you. I also think you should be able to chooose your partner.
> (no name)

> I think that computers would be quite good in schools and they would teach children how to type.
> (Harriet)

> I liked every part of this disc [Dragon World] except I think I would like it better if I went with somebody else - I were on my own.
> (Nicola)

> When you sit next to the computer it feels different than when you sit in front of it.
> (Mark)

> I learnt a lot about computers and wished that we had the computer until the end of the year.
> (Rachel)

The children at Somerfield School spent some time on a word-processing package, and although this was not formally part of the experiment, they also found this was worthwhile.

> It is a lot of fun writing a story on the printer - it comes out neater too ... You can delete with no messy crossing out which some people get confused with.
> (Nicola)

> Joanne and I wrote some more of our story but forgot to save it ... [next day] We wrote in the story that hadn't been saved.
> (Nicola)

> I enjoy writing stories on the computer better than on paper.
> (Joanne)

Finally, the children showed a fine sense of appreciation that they were the lucky ones who had been chosen to take part in this experiment, and no doubt were the object of many envious glances. They understood the value of what they were doing, looked to the future, and generally felt that their parents thoroughly approved.

My brother and sister thought we were lucky.
(Marnie)

A lot of people in classes that didn't have a computer thought we were lucky.
(Marcus)

I think computers are good to use because we will probably use them in the future.
(no name)

My parents said it was good we were getting to know computers, because we might use them later on in our life.
(Virginia)

General Teacher Comment

A round table discussion was held during the two-day meeting held at the Christchurch Teachers College on 2-3 December, 1986, and the following comments about the study were collected. They form a representative collection of views of teachers at all three schools about the way in which the study went, and reflect opinions expressed in their diaries.

Strengths of the programme - Introductory phase

It was really good. They think independently; it was good for discovery learning; they think logically. Good programs were *Flowers of Crystal* and *Dragon World*.

Telebook was particularly good; their spelling was much improved; their reading was improved - the poorer readers tried very hard ... *L* and *Flowers of Crystal* were too hard.

Flowers of Crystal and *Dragon World* and *L* were good. The children often worked at home on the *L* problems and demanded the opportunity to try their solutions the next day. They worked on their own at lunchtimes. They were fine on their own.

Dragon World went down well, but they got sick of it. The brighter children liked *Mazes* and *Colditz*. There didn't seem to be a correlation between computer experience and intelligence. I paired the children into 'computer haves' and 'have nots'. They liked the pairing. It wasn't always the brighter one who took the lead. Overall their self esteem seemed improved - especially the slower children.

In my class the children can choose their own groups. All the children had half an hour per two days. They were very enthusiastic to want to get on the computers; before school, at lunch time and after school. Games which were popular were *Flowers of Crystal*, which were very advanced for Std 4; they went for *Chess*, *Dragon World* and *L*. My Std 4s were all enthusiastic. They picked it up very quickly. I used peer tutoring. *Telebook* proved most worthwhile. I had an 8.30 a.m. to 4.00.

p.m. timetable, so there was no problem getting them on. They were very keen on *Blitz*. *Telebook* brought out an awareness of errors - a sense of achievement.

Difficulties noticed - Introductory phase

You can't turn the noise of the computer off!

We had a small classroom. The computers were at the back of the room. The children at the back were distracted. I got used to the noise but the other teachers who came in didn't. I ended blocking it out of maths and reading time because the noise was too disruptive.

The noise was too high. I ended up not allowing it between 9.00 and 10.30 and my reading period immediately after lunch. I insist on absolute quiet from the children - and other things! - at reading time. Sports periods and so on rather disrupted the computer use.

We had a really solid partition at the back - the computer was isolated. The noise level was still too high. In the end we shifted it out to another room (for the testing period only).

We had an ideal setup from the noise point of view. Later we used a withdrawal room and the computer was going all day. I sent the kids out for one hour sessions. We trained up resource students to help with the problems.

The computers were down the back on a bench. One of the screens was just about impossible to read and breakdowns were common - every time we moved them they wouldn't go again. Chalk dust was a problem.

The Teachers College technician fixed the plug.

Some of the children complained that there were insufficient notes for the games. They didn't know what to do.

Operation of the mathematics diagnostic program

The five weeks introductory work made the arithmetic bit very easy to administer. The children were so used to it.

In the maths program they were frustrated because it said to 'Push a number and they did and then it stopped.' [This was a reference to the experience at Somerfield School, already referred to, which led to the modification of the prompt 'Type a number' to 'Type your answer'.]

The problems we had were eliminated by the time you [the other schools] got it!

They pressed 'the wrong button' and it took them back to the beginning. [The children must have pushed BREAK, which wasn't disabled at this time.]

We had a hardware problem and our results weren't recorded.

Some of my children - the slower ones - agonised over it. They took up to an hour. They wrote all of the problems down.

They started to compare notes - 'What level did you get to? [It was agreed that since the exercise was supposed to be diagnostic, rather than achievement-based, the levels feedback should be removed.]

A teacher wish list

That the hardware would be more reliable.

A colour monitor would have been nice.

More opportunity for group discussions amongst all of us teachers so we could learn from each other's experiences.

The two days in-service at the beginning was invaluable.

I would like to have the feedback so I could use the results of the diagnostic test.

General Observations

Aside from the use of *Telebook* in a couple of classrooms, there was little attempt to use the introductory phase as part of the normal classroom programme. The introductory activities were a 'tack on', just to get the children used to the presence and use of the computer. All the teachers agreed, however, that the introductory experiences had been educationally worthwhile; they reported things like '...they were problem solving', '...it was good for logical thinking', '...they spent a lot of time discussing'.

Only within the open plan setting at Somerfield School, where children were used to working at activity tables, was the presence of the computer not considered in some way 'disruptive' of the normal routine. It may have been that the 'disruption' was considered as such because the activity was not perceived as a normal and necessary part of the work programme. However, the necessity of being able to control the sound levels on classroom courseware was reinforced.

6 CONCLUSIONS

From the first complete year of the experiment, some conclusions can been reached in relation to the main aims of the study. The more limited, informal replication of the research in 1987 has also made it possible to validate these findings with parallel results from the same schools in 1987.

Validity

The computer-based diagnostic programs compare quite favourably with the usual paper-and-pencil versions of the Seville Diagnostic Arithmetic tests, as modified for use in this exploratory survey, in the diagnosis of errors in the elementary operations of addition, subtraction and multiplication. In the major 1986 study, the mean number of errors detected by the programs, by the remedial teacher and by the class teachers involved, did not differ significantly, in any of the three subtests; the sample sizes were of course very small. If anything, the computer tended to detect more errors, and because of the way in which the programs were designed, they presented more examples to pupils on levels where they were experiencing difficulty, and thus were potentially more accurate then the usual paper-and-pencil versions, with their two items per level.

On a level-by-level basis, too, the computer versions of the tests showed a substantial match with the paper-and-pencil versions, particularly in addition and subtraction, and would lead to similar diagnoses of problems being experienced by Std 3 and Std 4 children in the various test objectives.

In the less well controlled 1987 replication the computer version of the test showed up more errors than the pencil and paper version. This suggested the importance of some guidance from the teacher at the outset to ensure that data entry procedures and conventions are fully understood by all children. Standard 2 children found difficulties with both the paper-and-pencil version and the computer version of the test, and accordingly it is not recommended for use with this age group.

Efficiency

The computer versions of the tests, on average, took somewhat less time to

administer, largely because all pupils entered the tests part way through, and those who were not finding too much difficulty moved upwards on alternate levels only, or in variable sized jumps related to the content of the items in the 1987 modifications. It is conceivable that with a group of less able children they could take *longer*, because of the built-in provision to increase the number of randomly-generated items presented to pupils at the point where they first begin to experience failure.

The computer tests did take longer *per level*, probably because of the initially unfamiliar nature of the interaction with the keyboard and screen, and the fact that it was necessary for some children to use paper and pencil as well, to write down such things as 'carrying' figures, before entering them. But the self-paced nature of the computer tests can be seen as a real advantage, however long they may have taken, because the teacher was not required to supervise the process. They score highly, therefore, on the grounds of efficiency.

Ease of Use

After some initial 'teething troubles', overcome by 'fine-tuning' the software, the programs appeared to be robust and easy for the children to use. Their general reaction to the exercise was very positive, and most of them were able to progress through the three programs, without undue boredom or frustration, and allow them to terminate normally, bearing a full cargo of diagnostic information held on disk for subsequent remediation by the classroom teacher. Some hardware faults caused problems at the beginning, but these did not persist once the causes were isolated.

Usability of Results

The computer diagnostic version scores very highly in this regard. One teacher not involved in the study, but who observed a presentation of the research, was overheard to remark, 'If I could get a sheet like that [the computer diagnostic output] for my class, it would be the most useful thing in 20 years'. While admittedly based on a diagnostic test compiled many years ago without the aid of modern Item Response Theory techniques, a diagnostic approach which automatically bypasses items or groups of items which a child finds easy, and offers an increased number of items for an objective on which a pupil is finding difficulty, along with full error printouts, has the potential to be a sharply focussed and very helpful classroom aid indeed.

REFERENCES

Anderson, C.J. (1918). The use of the Woody Scale for diagnostic purposes. *Elementary School Journal, 16*, 770-781.

Bennett, N., Desforges, C., Cockburn, A., and Wilkinson, B. (1984). *The Quality of Pupil Learning Experiences*. London: Lawrence Erlbaum.

Black, H.D. (1983). Introducing diagnostic assessment. *Programmed Learning and Educational Technology, 20*, 1, 58-63.

Brown, L.S., and Burton, R.R. (1978). Diagnostic models for procedural bugs in mathematical skills. *Cognitive Science, 2*, 155-192.

Choppin, B.H. (1983). *Extracting more information from multiple-choice tests: Analytic techniques for answer-until-correct mode*. (ED227175)

Connolly, A.J., Nachtman, W., and Pritchett, E.M. (1971). *Keymath diagnostic arithmetic test*. Circle Prince, Minnesota: American Guidance Service.

Denvir, Brenda and Brown, Margaret (1987). The feasibility of class administered diagnostic assessment in primary mathematics. *Educational Research, 29*, 2, 95-107.

Furlong, F. and Miller, W. (1978). DIAGNOSE: Computer-based reporting of criterion-referenced test results. *Educational Technology, 8*, 37-39.

Hart, K.M. Ed., (1981). *Children's Understanding of Mathematics 11-16*. London: John Murray.

McArthur, D.L. and Choppin, B.H. (1983). *Evaluating Diagnostic Hypotheses*. Research Report prepared for National Institute of Education. Center for the Study of Evaluation, California University, Los Angeles. (ED238933)

McArthur, D.L. and Choppin, B.H. (1984). Computerized diagnostic testing. *J. of Educational Measurement, 21*, 4, 391-397.

McBride, James R. (1985). Computerized adaptive testing. *Educational Leadership, 43*, 2, 25-28.

New Zealand Department of Education (1987). *Exploratory Studies in Educational Computing*. [Wellington]: Computers in Education Development Unit.

Signer, B. (1982). Math Doctor M.D. - Microcomputer adaptive diagnosis. *The Computing Teacher, 10*, 4, 16-18.

Smith, A.K. (Ed) (1987). *Mathematics Remediation Package*. Tasmania: Department of Education.

Tatsuoka, Kikumi and Birembaum, Menucha (1981). Effects of instructional backgrounds on test performances. *J. of Computer-Based Instruction, 8*, 1, 1-8.

Thomas, R.M. (1981). A model of diagnostic evaluation. In A. Levy and D. Nevo Eds., *Evaluation Roles in Education*. London: Gordon and Breach.

Weiss, D.J. Ed., (1983). *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. New York: Academic Press.

Weiss, D.J. (1985). Adaptive testing by computer. *J. of Consulting and Clinical Psychology, 53*, 6, 774-89.

APPENDIX A

SEVILLE DIAGNOSTIC ARITHMETIC TESTS

TRIAL VERSION

SEVILLE DIAGNOSTIC TEST OF COMPUTATIONAL SKILLS - ADDITION.

Type 1. Extensions of basic facts, within ten.

$\begin{array}{r} 21 \\ 6 \\ \hline \end{array}$ $\begin{array}{r} 34 \\ 2 \\ \hline \end{array}$

Type 2. Extensions of basic facts, within ten.

$\begin{array}{r} 3 \\ 23 \\ \hline \end{array}$ $\begin{array}{r} 4 \\ 12 \\ \hline \end{array}$

Type 3. Extensions, higher decades.

$\begin{array}{r} 81 \\ 6 \\ \hline \end{array}$ $\begin{array}{r} 73 \\ 5 \\ \hline \end{array}$

Type 4. Extensions, higher decades.

$\begin{array}{r} 1 \\ 85 \\ \hline \end{array}$ $\begin{array}{r} 2 \\ 64 \\ \hline \end{array}$

Type 5. Two, 2 digit addends, no renaming.

$\begin{array}{r} 22 \\ 12 \\ \hline \end{array}$ $\begin{array}{r} 34 \\ 32 \\ \hline \end{array}$

Type 6. One addend a multiple of ten.

$\begin{array}{r} 10 \\ 47 \\ \hline \end{array}$ $\begin{array}{r} 40 \\ 24 \\ \hline \end{array}$

Type 7. Both addends multiples of ten.

$\begin{array}{r} 60 \\ 20 \\ \hline \end{array}$ $\begin{array}{r} 20 \\ 50 \\ \hline \end{array}$

Type 8. Extensions of basic facts, bridging ten

$\begin{array}{r} 14 \\ 8 \\ \hline \end{array}$ $\begin{array}{r} 16 \\ 8 \\ \hline \end{array}$

Type 9. Exten'ns of basic facts, bridging ten.

$\begin{array}{r} 6 \\ 16 \\ \hline \end{array}$ $\begin{array}{r} 6 \\ 19 \\ \hline \end{array}$

Type 10. Exten'ns, bridging, higher decades.

$\begin{array}{r} 78 \\ 8 \\ \hline \end{array}$ $\begin{array}{r} 83 \\ 9 \\ \hline \end{array}$

Type 11. Exten'ns, bridging, higher decades.

$$\begin{array}{r} 4 \\ 28 \\ \hline \end{array} \qquad \begin{array}{r} 5 \\ 76 \\ \hline \end{array}$$

Type 12. Two 2d addends, renaming from ones.

$$\begin{array}{r} 24 \\ 28 \\ \hline \end{array} \qquad \begin{array}{r} 38 \\ 27 \\ \hline \end{array}$$

Type 13. 2d addends, 3d sum, renaming from tens.

$$\begin{array}{r} 51 \\ 63 \\ \hline \end{array} \qquad \begin{array}{r} 42 \\ 85 \\ \hline \end{array}$$

Type 14. 2d addends, renaming ones and tens.

$$\begin{array}{r} 95 \\ 86 \\ \hline \end{array} \qquad \begin{array}{r} 76 \\ 78 \\ \hline \end{array}$$

Type 15. Three addends bridging ten, equat'n.

$$(4 + 1) + 8 =$$

$$(5 + 3) + 4 =$$

Type 16. Three addends bridging ten, vertical.

$$\begin{array}{r} 5 \\ 2 \\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 6 \\ 3 \\ 5 \\ \hline \end{array}$$

Type 17. Three addends each 2d, renaming.

$$\begin{array}{r} 26 \\ 43 \\ 52 \\ \hline \end{array} \qquad \begin{array}{r} 37 \\ 24 \\ 86 \\ \hline \end{array}$$

Type 18. Three 2d addends, zero diffs.

$$\begin{array}{r} 15 \\ 47 \\ 30 \\ \hline \end{array} \qquad \begin{array}{r} 26 \\ 54 \\ 70 \\ \hline \end{array}$$

Type 19. Four addends, one and two digits.

$$\begin{array}{r} 26 \\ 3 \\ 18 \\ 2 \\ \hline \end{array} \qquad \begin{array}{r} 8 \\ 36 \\ 4 \\ 16 \\ \hline \end{array}$$

Type 20. 2d addends, sum multiple of ten.

$$\begin{array}{r} 28 \\ 62 \\ \hline \end{array} \qquad \begin{array}{r} 34 \\ 36 \\ \hline \end{array}$$

Type 21. 2d addends, sum one hundred.

$$\begin{array}{r} 21 \\ \underline{79} \end{array} \qquad \begin{array}{r} 37 \\ \underline{63} \end{array}$$

Type 22. Two 3d addends, no renaming.

$$\begin{array}{r} 123 \\ \underline{711} \end{array} \qquad \begin{array}{r} 234 \\ \underline{125} \end{array}$$

Type 23. A 3d and a 2d addend, no renaming.

$$\begin{array}{r} 421 \\ \underline{36} \end{array} \qquad \begin{array}{r} 832 \\ \underline{24} \end{array}$$

Type 24. A 3d and a 2d addend, no renaming.

$$\begin{array}{r} 18 \\ \underline{311} \end{array} \qquad \begin{array}{r} 34 \\ \underline{425} \end{array}$$

Type 25. 3d addends,4d sum,renaming from hund's.

$$\begin{array}{r} 432 \\ \underline{841} \end{array} \qquad \begin{array}{r} 314 \\ \underline{864} \end{array}$$

Type 26. 3d addends,renaming ones and tens.

$$\begin{array}{r} 348 \\ \underline{478} \end{array} \qquad \begin{array}{r} 627 \\ \underline{297} \end{array}$$

Type 27. Four addends,1,2 and 3d mixture.

$$\begin{array}{r} 23 \\ 426 \\ 9 \\ \underline{13} \end{array} \qquad \begin{array}{r} 2 \\ 27 \\ 384 \\ \underline{16} \end{array}$$

Type 28. Zero diffs in sum.

$$\begin{array}{r} 528 \\ \underline{492} \end{array} \qquad \begin{array}{r} 381 \\ \underline{689} \end{array}$$

Type 29. Three addends, zeros in tens col.

$$\begin{array}{r} 203 \\ 204 \\ \underline{305} \end{array} \qquad \begin{array}{r} 408 \\ 104 \\ \underline{406} \end{array}$$

Type 30. Four 4d addends, renaming all cols.

$$\begin{array}{r} 4681 \\ 3214 \\ 6859 \\ \underline{3472} \end{array} \qquad \begin{array}{r} 9652 \\ 7381 \\ 2976 \\ \underline{8322} \end{array}$$

SEVILLE DIAGNOSTIC TEST OF COMPUTATIONAL SKILLS - SUBTRACTION.

Type 1. Single column, no adjustment. $\begin{array}{r} 35 \\ -\ 3 \\ \hline \end{array}$ $\begin{array}{r} 48 \\ -\ 6 \\ \hline \end{array}$	Type 2. Tens and ones, no adjustment. $\begin{array}{r} 48 \\ -23 \\ \hline \end{array}$ $\begin{array}{r} 56 \\ -43 \\ \hline \end{array}$
Type 3. Tens digits the same, no adjustment. $\begin{array}{r} 19 \\ -11 \\ \hline \end{array}$ $\begin{array}{r} 37 \\ -34 \\ \hline \end{array}$	Type 4. Whole tens from whole tens, no adj. $\begin{array}{r} 46 \\ -10 \\ \hline \end{array}$ $\begin{array}{r} 39 \\ -20 \\ \hline \end{array}$
Type 5. Known addend a whole ten, no adj. $\begin{array}{r} 40 \\ -20 \\ \hline \end{array}$ $\begin{array}{r} 50 \\ -40 \\ \hline \end{array}$	Type 6. Zero in ones answer, no adj. $\begin{array}{r} 37 \\ -27 \\ \hline \end{array}$ $\begin{array}{r} 45 \\ -25 \\ \hline \end{array}$
Type 7. Adjustment, one digit known addend. $\begin{array}{r} 44 \\ -\ 7 \\ \hline \end{array}$ $\begin{array}{r} 32 \\ -\ 6 \\ \hline \end{array}$	Type 8. Adjustment, two digit known addend. $\begin{array}{r} 84 \\ -27 \\ \hline \end{array}$ $\begin{array}{r} 64 \\ -28 \\ \hline \end{array}$
Type 9. Adjustment, zero answer in tens. $\begin{array}{r} 41 \\ -39 \\ \hline \end{array}$ $\begin{array}{r} 72 \\ -66 \\ \hline \end{array}$	Type 10. Adj., zero difficulty in ones. $\begin{array}{r} 50 \\ -15 \\ \hline \end{array}$ $\begin{array}{r} 60 \\ -46 \\ \hline \end{array}$

Type 11. Adjustment 0 - 9 in ones. 30 60 -19 -29	Type 12. Adjustment 0 - 1 in ones. 70 40 -21 -11
Type 13. Three columns, no adjustment. 445 684 -231 -463	Type 14. 2d known addend, adj., ones only. 452 648 - 28 - 19
Type 15. 2d known addend, adj., tens only. 485 748 - 92 - 75	Type 16. 2d known addend, adj., ones and tens. 645 834 - 66 - 68
Type 17. Three columns,adj.,ones and tens. 642 856 -164 -259	Type 18. Three col,adj.,in ones, 1d answer. 456 672 -448 -667
Type 19. Adj., in ones,zero diff in tens. 354 554 -207 -108	Type 20. Adj., in ones ,zero diff in tens sum. 406 603 -128 -247

Type 21. Adj., in ones, zeros in tens.

$$\begin{array}{r} 406 \\ -208 \\ \hline \end{array} \qquad \begin{array}{r} 602 \\ -309 \\ \hline \end{array}$$

Type 22. Adj., in ones, zero in tens ans.

$$\begin{array}{r} 452 \\ -246 \\ \hline \end{array} \qquad \begin{array}{r} 784 \\ -377 \\ \hline \end{array}$$

Type 23. Adj.,ones and tens, 9 in tens k.a.

$$\begin{array}{r} 464 \\ -197 \\ \hline \end{array} \qquad \begin{array}{r} 846 \\ -298 \\ \hline \end{array}$$

Type 24. Adj.,ones and tens,0 - 9 in tens.

$$\begin{array}{r} 704 \\ -397 \\ \hline \end{array} \qquad \begin{array}{r} 402 \\ -196 \\ \hline \end{array}$$

Type 25. 1d known addend, with adjustment.

$$\begin{array}{r} 462 \\ -\ \ 6 \\ \hline \end{array} \qquad \begin{array}{r} 372 \\ -\ \ 9 \\ \hline \end{array}$$

Type 26. 2d k.a.,zero answers in ones, tens.

$$\begin{array}{r} 229 \\ -\ 29 \\ \hline \end{array} \qquad \begin{array}{r} 746 \\ -\ 46 \\ \hline \end{array}$$

Type 27. Zero answers in ones and hundreds.

$$\begin{array}{r} 640 \\ -610 \\ \hline \end{array} \qquad \begin{array}{r} 370 \\ -320 \\ \hline \end{array}$$

Type 28. 0 -9 diffs in ones and tens.

$$\begin{array}{r} 600 \\ -299 \\ \hline \end{array} \qquad \begin{array}{r} 500 \\ -399 \\ \hline \end{array}$$

Type 29. Four cols, adj., in ones, tens, hunds.

$$\begin{array}{r} 4216 \\ -2769 \\ \hline \end{array} \qquad \begin{array}{r} 9253 \\ -6469 \\ \hline \end{array}$$

Type 30. Double zero diffs.

$$\begin{array}{r} 6214 \\ -1008 \\ \hline \end{array} \qquad \begin{array}{r} 8614 \\ -4009 \\ \hline \end{array}$$

SEVILLE DIAGNOSTIC TEST OF COMPUTATIONAL SKILLS - SIMPLE MULTIPLICATION.

Type 1. 2d in multiplicand, no renaming.

$$\begin{array}{r} 42 \\ \times\ 2 \\ \hline \end{array} \qquad \begin{array}{r} 23 \\ \times\ 3 \\ \hline \end{array}$$

Type 2. 3d in multiplicand, no renaming.

$$\begin{array}{r} 321 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 313 \\ \times\ 3 \\ \hline \end{array}$$

Type 3. 2d in mul'nd, 3d in product,no ren'g.

$$\begin{array}{r} 41 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 62 \\ \times\ 4 \\ \hline \end{array}$$

Type 4. Zero in ones of 2d mul'nd,no ren'g.

$$\begin{array}{r} 30 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 20 \\ \times\ 3 \\ \hline \end{array}$$

Type 5. Zero in ones of 3d mul'nd,no ren'g.

$$\begin{array}{r} 320 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 120 \\ \times\ 4 \\ \hline \end{array}$$

Type 6. Zero in tens of 3d mul'nd,no ren'g.

$$\begin{array}{r} 202 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 402 \\ \times\ 2 \\ \hline \end{array}$$

Type 7. Double zero in mul'nd, no renaming.

$$\begin{array}{r} 300 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 200 \\ \times\ 4 \\ \hline \end{array}$$

Type 8. Zero in tens col of product, no ren'g.

$$\begin{array}{r} 41 \\ \times\ 5 \\ \hline \end{array} \qquad \begin{array}{r} 52 \\ \times\ 2 \\ \hline \end{array}$$

Type 9. 2d multiplicand, renaming from ones.

$$\begin{array}{r} 13 \\ \times\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 15 \\ \times\ 5 \\ \hline \end{array}$$

Type 10. 2d multiplicand, renaming from ones.

$$\begin{array}{r} 25 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 28 \\ \times\ 3 \\ \hline \end{array}$$

Type 11. 2d in mul'nd, 3d in product,no ren'g.

$$\begin{array}{r} 48 \\ \times\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 67 \\ \times\ 5 \\ \hline \end{array}$$

Type 12. 3d multiplicand, renaming from tens.

$$\begin{array}{r} 271 \\ \times\ 2 \\ \hline \end{array} \qquad \begin{array}{r} 253 \\ \times\ 3 \\ \hline \end{array}$$

Type 13. Renaming from ones with zero tens.

$$\begin{array}{r} 105 \\ \times\ 5 \\ \hline \end{array} \qquad \begin{array}{r} 208 \\ \times\ 4 \\ \hline \end{array}$$

Type 14. 3d mul'nd, 4d product, no ren'g.

$$\begin{array}{r} 621 \\ \times\ 4 \\ \hline \end{array} \qquad \begin{array}{r} 312 \\ \times\ 4 \\ \hline \end{array}$$

Type 15. 4d mul'nd, doble zeros.

$$\begin{array}{r} 3002 \\ \times\ 3 \\ \hline \end{array} \qquad \begin{array}{r} 2003 \\ \times\ 2 \\ \hline \end{array}$$

Type 16. 4d mul'nd, ren'g from ones   hunds.

$$\begin{array}{r} 1213 \\ \times\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 2418 \\ \times\ 4 \\ \hline \end{array}$$

Type 17. Ren'g from 2 places, within tens.

$$\begin{array}{r} 583 \\ \times\ 6 \\ \hline \end{array} \qquad \begin{array}{r} 754 \\ \times\ 5 \\ \hline \end{array}$$

Type 18. Ren'g from 2 places, bridging tens.

$$\begin{array}{r} 146 \\ \times\ 9 \\ \hline \end{array} \qquad \begin{array}{r} 689 \\ \times\ 6 \\ \hline \end{array}$$

Type 22. Multiplier multiples of 10,no ren'g.

$$\begin{array}{r} 46 \\ \times 10 \\ \hline \end{array} \qquad \begin{array}{r} 13 \\ \times 30 \\ \hline \end{array}$$

Type 23. Multiplier multiple of 10,renaming.

$$\begin{array}{r} 38 \\ \times 40 \\ \hline \end{array} \qquad \begin{array}{r} 47 \\ \times 60 \\ \hline \end{array}$$

Type 24. As for type 23, plus zero diff.

$$\begin{array}{r} 35 \\ \times 40 \\ \hline \end{array} \qquad \begin{array}{r} 75 \\ \times 80 \\ \hline \end{array}$$

Type 25. 4d mul'nd, ren'g with bridging tens.

$$\begin{array}{r} 3368 \\ \times \quad 3 \\ \hline \end{array} \qquad \begin{array}{r} 2769 \\ \times \quad 8 \\ \hline \end{array}$$

Type 26. 4d mul'nd, zero in ones column.

$$\begin{array}{r} 6740 \\ \times \quad 6 \\ \hline \end{array} \qquad \begin{array}{r} 4380 \\ \times \quad 5 \\ \hline \end{array}$$

Type 27. 4d mul'nd, zero in tens column.

$$\begin{array}{r} 7606 \\ \times \quad 4 \\ \hline \end{array} \qquad \begin{array}{r} 8406 \\ \times \quad 3 \\ \hline \end{array}$$

Type 28. 4d mul'nd, zero in hundreds column.

$$\begin{array}{r} 4069 \\ \times \quad 4 \\ \hline \end{array} \qquad \begin{array}{r} 3028 \\ \times \quad 3 \\ \hline \end{array}$$

SEVILLE DIAGNOSTIC ARITHMETIC TESTS

REVISED VERSION

ADDTEST - PROGRESSION LEVELS.

1. Basic facts, sentence form, sums less than 10	$4 + 3 = \square$
2. Basic facts, vertical form, sums less than 10	$\begin{array}{r} 5 \\ +\ 3 \\ \hline \end{array}$
3. Basic facts, sentence form, bridging ten, sums < 19	$5 + 7 = \square$
4. Basic facts, vertical form, bridging ten, sums < 19	$\begin{array}{r} 6 \\ +\ 8 \\ \hline \end{array}$
5. Extensions of basic facts, sentence form, no bridging, sums less than 50	$22 + 5 = \square$
6. Extensions of basic facts, vertical form, no bridging, sums less than 50	$\begin{array}{r} 6 \\ +\ 31 \\ \hline \end{array}$
7. Extensions of basic facts, sentence form, no bridging, sums less than 100	$63 + 2 = \square$
8. Extensions of basic facts, vertical form, no bridging, sums less than 100	$\begin{array}{r} 4 \\ +\ 74 \\ \hline \end{array}$
9. Extensions of basic facts, sentence form, bridging ten, sums less than 50	$14 + 8 = \square$
10. Extensions of basic facts, vertical form, bridging ten, sums less than 50	$\begin{array}{r} 7 \\ +\ 36 \\ \hline \end{array}$
11. Extensions of basic facts, sentence form, bridging ten, sums less than 100	$75 + 7 = \square$
12. Extensions of basic facts, vertical form, bridging ten, sums less than 100	$\begin{array}{r} 4 \\ +\ 59 \\ \hline \end{array}$
13. Three digits, vertical form	$\begin{array}{r} 8 \\ 4 \\ +\ 9 \\ \hline \end{array}$
14. Three digits, sentence form	$4 + 5 + 6 = \square$
15. Whole tens, sums less than 100	$\begin{array}{r} 40 \\ +\ 30 \\ \hline \end{array}$

16.	Whole tens, sums greater than 100	40 + 80
17.	Tens and ones, no renaming	42 + 25
18.	Tens and ones, renaming ones	47 + 25
19.	Tens and ones, renaming tens	53 + 62
20.	Tens and ones, renaming ones and tens	47 + 86
21.	Three columns, no renaming	245 + 132
22.	Three columns, renaming ones	247 + 128
23.	Three columns, renaming ones and tens	236 + 198
24.	Three columns, renaming all three	764 + 398
25.	Three addends, each three columns, with renaming	724 435 + 146
26.	Three addends, columns with empty spaces	241 26 + 102

SUBTEST - PROGRESSION LEVELS.

1.	Basic facts, vertical form, sum less than 10	$\begin{array}{r} 9 \\ -\ 3 \\ \hline \end{array}$
2.	Basic facts, sum less than 20, with bridging	$\begin{array}{r} 15 \\ -\ 8 \\ \hline \end{array}$
3.	Two digit sum, one digit known addend, no adjustment	$\begin{array}{r} 39 \\ -\ 7 \\ \hline \end{array}$
4.	Tens and ones, no adjustment	$\begin{array}{r} 38 \\ -\ 26 \\ \hline \end{array}$
5.	Tens and ones, tens digits the same, no adjustment	$\begin{array}{r} 27 \\ -\ 24 \\ \hline \end{array}$
6.	Tens and ones, known addend a whole ten	$\begin{array}{r} 69 \\ -\ 30 \\ \hline \end{array}$
7.	Sum and known addend both whole tens	$\begin{array}{r} 80 \\ -\ 20 \\ \hline \end{array}$
8.	Tens and ones, zero answer in ones column	$\begin{array}{r} 71 \\ -\ 41 \\ \hline \end{array}$
9.	Two digit sum, one digit known addend, adjustment	$\begin{array}{r} 35 \\ -\ 9 \\ \hline \end{array}$
10.	Tens and ones with adjustment	$\begin{array}{r} 43 \\ -\ 18 \\ \hline \end{array}$
11.	Tens and ones, with adjustment, zero answer in tens	$\begin{array}{r} 35 \\ -\ 27 \\ \hline \end{array}$
12.	Tens and ones, sum a whole ten	$\begin{array}{r} 50 \\ -\ 34 \\ \hline \end{array}$
13.	Three digit sum, one digit known addend, no adj.	$\begin{array}{r} 129 \\ -\ 4 \\ \hline \end{array}$
14.	Three digit sum, two digit known addend, no adj.	$\begin{array}{r} 386 \\ -\ 52 \\ \hline \end{array}$

15. Three columns, no adjustment

$$\begin{array}{r} 569 \\ -\,344 \\ \hline \end{array}$$

16. Three digit sum, one digit known addend, with adj.

$$\begin{array}{r} 242 \\ -\quad 6 \\ \hline \end{array}$$

17. Three digit sum, two digit known addend, with adjustment in ones only

$$\begin{array}{r} 255 \\ -\;\; 36 \\ \hline \end{array}$$

18. Three columns with adjustment in ones and tens

$$\begin{array}{r} 531 \\ -\,276 \\ \hline \end{array}$$

19. Three columns with adj., zero difficulty in sum

$$\begin{array}{r} 302 \\ -\,185 \\ \hline \end{array}$$

20. Three columns with adj., sum a whole hundred

$$\begin{array}{r} 400 \\ -\,276 \\ \hline \end{array}$$

MULTTEST - PROGRESSION LEVELS.

1.	Basic facts, sentence form, first factor 2,3,4,5	$4 \times 7 =$
2.	Basic facts, sentence form, first factor 6,7,8,9	$6 \times 7 =$
3.	Basic facts, vertical form, first factor 2,3,4,5	$\begin{array}{r} 5 \\ \times\ 3 \\ \hline \end{array}$
4.	Basic facts, vertical form, first factor 6,7,8,9	$\begin{array}{r} 7 \\ \times\ 8 \\ \hline \end{array}$
5.	Basic facts, vertical form, one factor zero	$\begin{array}{r} 0 \\ \times\ 5 \\ \hline \end{array}$ $\begin{array}{r} 5 \\ \times\ 0 \\ \hline \end{array}$
6.	Tens and ones, no renaming	$\begin{array}{r} 42 \\ \times\ 2 \\ \hline \end{array}$
7.	Whole tens, no renaming	$\begin{array}{r} 30 \\ \times\ 2 \\ \hline \end{array}$
8.	Tens and ones, zero in ones, renaming tens	$\begin{array}{r} 40 \\ \times\ 3 \\ \hline \end{array}$
9.	Tens and ones, renaming ones	$\begin{array}{r} 25 \\ \times\ 3 \\ \hline \end{array}$
10.	Tens and ones, renaming tens and ones, no bridging	$\begin{array}{r} 36 \\ \times\ 4 \\ \hline \end{array}$
11.	Tens and ones, with renaming and bridging	$\begin{array}{r} 45 \\ \times\ 7 \\ \hline \end{array}$
12.	Three columns, no renaming	$\begin{array}{r} 321 \\ \times\ 3 \\ \hline \end{array}$
13.	Three columns, no renaming, zero in ones	$\begin{array}{r} 430 \\ \times\ 2 \\ \hline \end{array}$
14.	Three columns, no renaming, zero in tens	$\begin{array}{r} 203 \\ \times\ 3 \\ \hline \end{array}$

15.	Three columns, no renaming, zero in tens and ones	$\begin{array}{r} 400 \\ \times\ 2 \\ \hline \end{array}$
16.	Three columns, renaming ones and tens, no bridging	$\begin{array}{r} 275 \\ \times\ 3 \\ \hline \end{array}$
17.	Three columns, renaming ones and tens, bridging	$\begin{array}{r} 475 \\ \times\ 7 \\ \hline \end{array}$
18.	Three columns, renaming, zero tens in factor	$\begin{array}{r} 205 \\ \times\ 3 \\ \hline \end{array}$
19.	Three columns, renaming, zero tens in product	$\begin{array}{r} 134 \\ \times\ 6 \\ \hline \end{array}$
20.	Three columns, renaming, zero tens and zero hundreds in product	$\begin{array}{r} 335 \\ \times\ 3 \\ \hline \end{array}$
21.	Four columns	$\begin{array}{r} 4631 \\ \times\ 5 \\ \hline \end{array}$

ADDTEST. 1

1.	4 + 5 = □	3 + 4 = □	7 + 2 = □	5 + 3 = □
2.	2 + 5 ____	6 + 3 ____	4 + 2 ____	4 + 4 ____
3.	6 + 7 = □	5 + 9 = □	8 + 3 = □	7 + 9 = □
4.	7 + 8 ____	9 + 6 ____	4 + 8 ____	8 + 6 ____
5.	23 + 5 = □	32 + 6 = □	16 + 3 = □	45 + 4 = □
6.	26 + 3 ____	2 + 33 ____	45 + 2 ____	4 + 13 ____
7.	64 + 3 = □	52 + 4 = □	76 + 2 = □	83 + 5 = □
8.	73 + 6 ____	2 + 57 ____	65 + 4 ____	7 + 91 ____
9.	14 + 8 = □	37 + 6 = □	25 + 9 = □	16 + 8 = □

2

10.	2 8 + 5 ____	8 + 3 7 ____	1 7 + 5 ____	3 + 2 8 ____
11.	77 + 5 = ☐	48 + 6 = ☐	88 + 4 = ☐	58 + 7 = ☐
12.	5 9 + 4 ____	8 + 6 3 ____	7 6 + 9 ____	8 + 8 5 ____
13.	3 5 + 4 ____	4 7 + 3 ____	6 7 + 8 ____	9 8 + 7 ____
14.	3 + 4 + 6 = ☐	6 + 8 + 5 = ☐	5 + 4 + 7 = ☐	8 + 5 + 9 = ☐
15.	2 0 + 4 0 ____	3 0 + 5 0 ____	6 0 + 2 0 ____	5 0 + 4 0 ____
16.	4 0 + 8 0 ____	6 0 + 5 0 ____	7 0 + 8 0 ____	9 0 + 8 0 ____
17.	3 5 + 5 1 ____	2 4 + 3 5 ____	1 6 + 5 2 ____	4 2 + 2 5 ____

3

18.	36 + 16	27 + 48	47 + 25	56 + 39
19.	52 + 64	71 + 56	63 + 74	45 + 73
20.	43 + 69	57 + 83	65 + 87	37 + 68
21.	245 + 132	327 + 521	236 + 462	412 + 157
22.	254 + 128	375 + 417	138 + 527	426 + 139
23.	467 + 246	356 + 178	239 + 387	147 + 458
24.	597 + 644	386 + 664	405 + 597	752 + 248
25.	724 156 + 349	608 497 + 610	75 380 + 609	937 21 + 62

73

1

SUBTEST.

1.	9 - 4	7 - 5	8 - 3	9 - 6
2.	14 - 6	16 - 9	15 - 7	17 - 8
3.	47 - 4	38 - 5	29 - 7	36 - 2
4.	56 - 24	67 - 43	79 - 65	87 - 35
5.	37 - 33	46 - 42	59 - 54	28 - 25
6.	47 - 20	53 - 40	65 - 40	72 - 30
7.	80 - 40	60 - 10	70 - 20	90 - 60
8.	72 - 52	68 - 38	53 - 23	86 - 66

2

9.	23 - 7	35 - 9	44 - 8	51 - 3
10.	32 - 16	43 - 26	35 - 18	47 - 19
11.	35 - 27	46 - 38	24 - 18	37 - 29
12.	40 - 27	50 - 35	60 - 24	70 - 46
13.	136 - 2	147 - 5	235 - 4	328 - 3
14.	175 - 52	386 - 34	467 - 42	269 - 57
15.	457 - 145	368 - 143	586 - 351	479 - 264
16.	242 - 6	156 - 7	361 - 9	425 - 8

					3
17.	$\begin{array}{r} 163 \\ -\ \ 37 \\ \hline \end{array}$	$\begin{array}{r} 245 \\ -\ \ 17 \\ \hline \end{array}$	$\begin{array}{r} 174 \\ -\ \ 58 \\ \hline \end{array}$	$\begin{array}{r} 356 \\ -\ \ 29 \\ \hline \end{array}$	
18.	$\begin{array}{r} 531 \\ -276 \\ \hline \end{array}$	$\begin{array}{r} 742 \\ -357 \\ \hline \end{array}$	$\begin{array}{r} 635 \\ -139 \\ \hline \end{array}$	$\begin{array}{r} 724 \\ -486 \\ \hline \end{array}$	
19.	$\begin{array}{r} 302 \\ -185 \\ \hline \end{array}$	$\begin{array}{r} 506 \\ -249 \\ \hline \end{array}$	$\begin{array}{r} 401 \\ -368 \\ \hline \end{array}$	$\begin{array}{r} 716 \\ -647 \\ \hline \end{array}$	
20	$\begin{array}{r} 400 \\ -276 \\ \hline \end{array}$	$\begin{array}{r} 500 \\ -415 \\ \hline \end{array}$	$\begin{array}{r} 600 \\ -493 \\ \hline \end{array}$	$\begin{array}{r} 300 \\ -234 \\ \hline \end{array}$	

MULTEST. 1

1.	3 x 6 = ☐	4 x 7 = ☐	2 x 9 = ☐	3 x 8 = ☐
2.	6 x 7 = ☐	8 x 5 = ☐	7 x 9 = ☐	9 x 6 = ☐
3.	4 x 3 ____	6 x 4 ____	5 x 3 ____	7 x 3 ____
4.	6 x 8 ____	5 x 7 ____	7 x 8 ____	9 x 8 ____
5.	0 x 4 ____	3 x 0 ____	0 x 7 ____	6 x 0 ____
6.	1 3 x 3 ____	4 2 x 2 ____	2 3 x 2 ____	1 2 x 4 ____
7.	2 0 x 4 ____	3 0 x 3 ____	1 0 x 7 ____	2 0 x 3 ____
8.	4 0 x 4 ____	5 0 x 6 ____	3 0 x 6 ____	2 0 x 7 ____

2

9.	26 x 3	17 x 4	39 x 2	18 x 4
10.	36 x 4	57 x 3	36 x 7	48 x 6
11.	37 x 6	75 x 7	28 x 8	37 x 9
12.	321 x 3	143 x 2	212 x 4	132 x 3
13.	340 x 2	120 x 4	210 x 3	410 x 2
14.	203 x 3	102 x 4	101 x 7	304 x 2
15.	200 x 4	300 x 3	400 x 2	300 x 2
16.	275 x 3	154 x 6	238 x 4	187 x 5

3

17.	475 × 7	426 × 8	837 × 6	347 × 9
18.	307 × 5	406 × 4	504 × 6	708 × 7
19.	234 × 6	635 × 3	167 × 9	643 × 7
20.	335 × 3	223 × 9	667 × 3	334 × 9
21.	4631 × 5	2543 × 7	6089 × 4	9006 × 8

APPENDIX B

SCATTERPLOTS OF TOTAL NUMBER OF ERRORS DETECTED, BY TESTER

1986 RESULTS

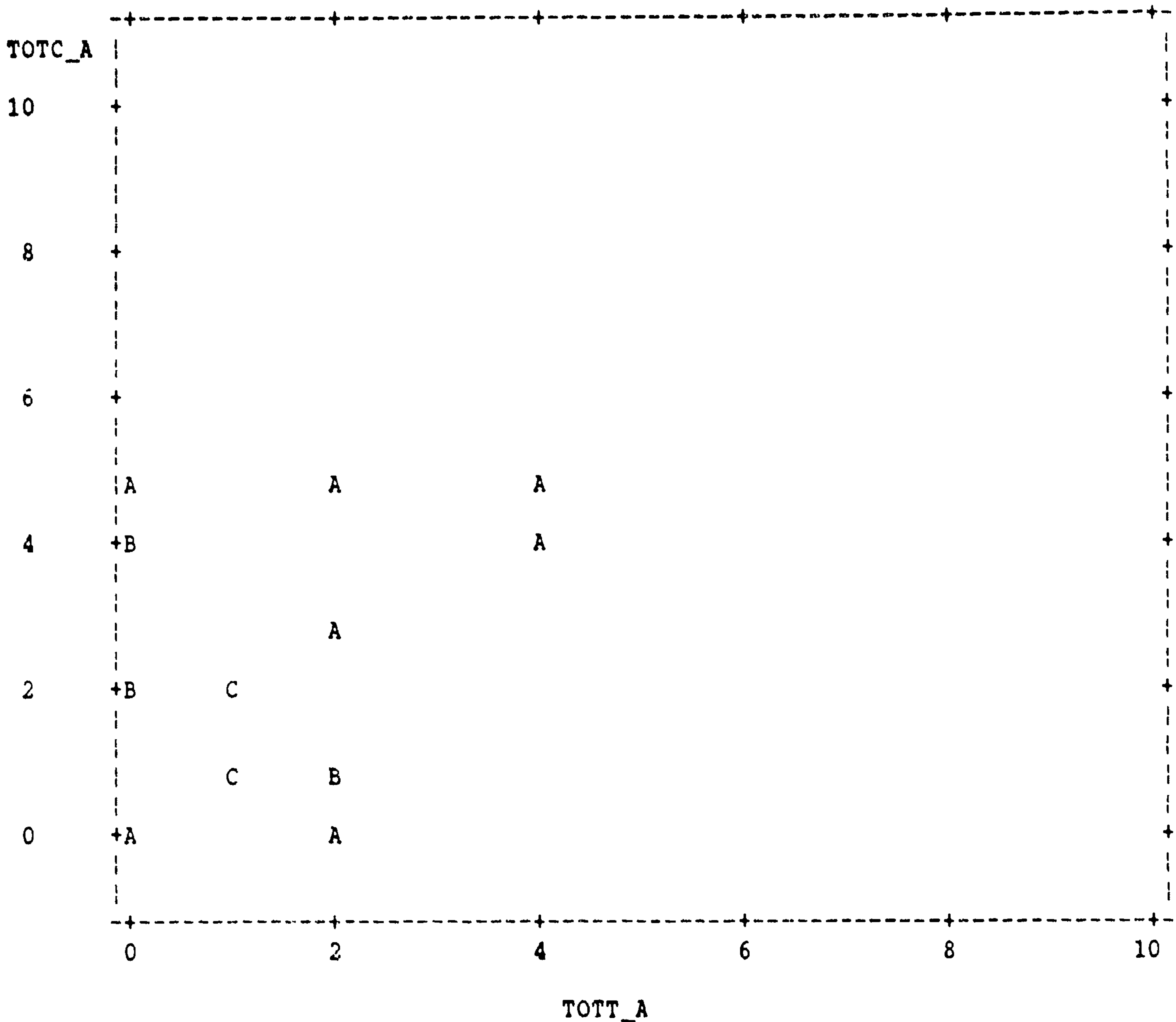
REMEDIAL TEACHER

```
               Plot of TOTC_A*TOTT_A    Legend:  A = 1 obs, B = 2 obs, etc.

          -+-------------+-------------+-------------+-------------+-------------+-
TOTC_A    |                                                                       |
          |                                                                       |
10        +                                                                       +
          |                                                                       |
          |                                                                       |
          |                                                                       |
          |                                                                       |
 8        +                                                                       +
          |                                                                       |
          |                                                                       |
          |                                                                       |
          |                                                                       |
 6        +                                                                       +
          |                                                                       |
          |                                                                       |
          |A                                                                      |
          |                                                                       |
 4        +A             A             A             B                            +
          |                                                                       |
          |                                                                       |
          |A      A                    A             A                            |
          |                                                                       |
 2        +B      A      A                                                        +
          |                                                                       |
          |                                                                       |
          |B      A                                                               |
          |                                                                       |
 0        +B      A                                                               +
          |                                                                       |
          |                                                                       |
          -+-------------+-------------+-------------+-------------+-------------+-
           0             2             4             6             8            10

                                       TOTT_A
```

KEY: TOTC_A Addition error detected by computer
TOTT_A Addition error detected by tester

CLASSROOM TEACHER

Plot of TOTC_A*TOTT_A   Legend: A = 1 obs, B = 2 obs, etc.

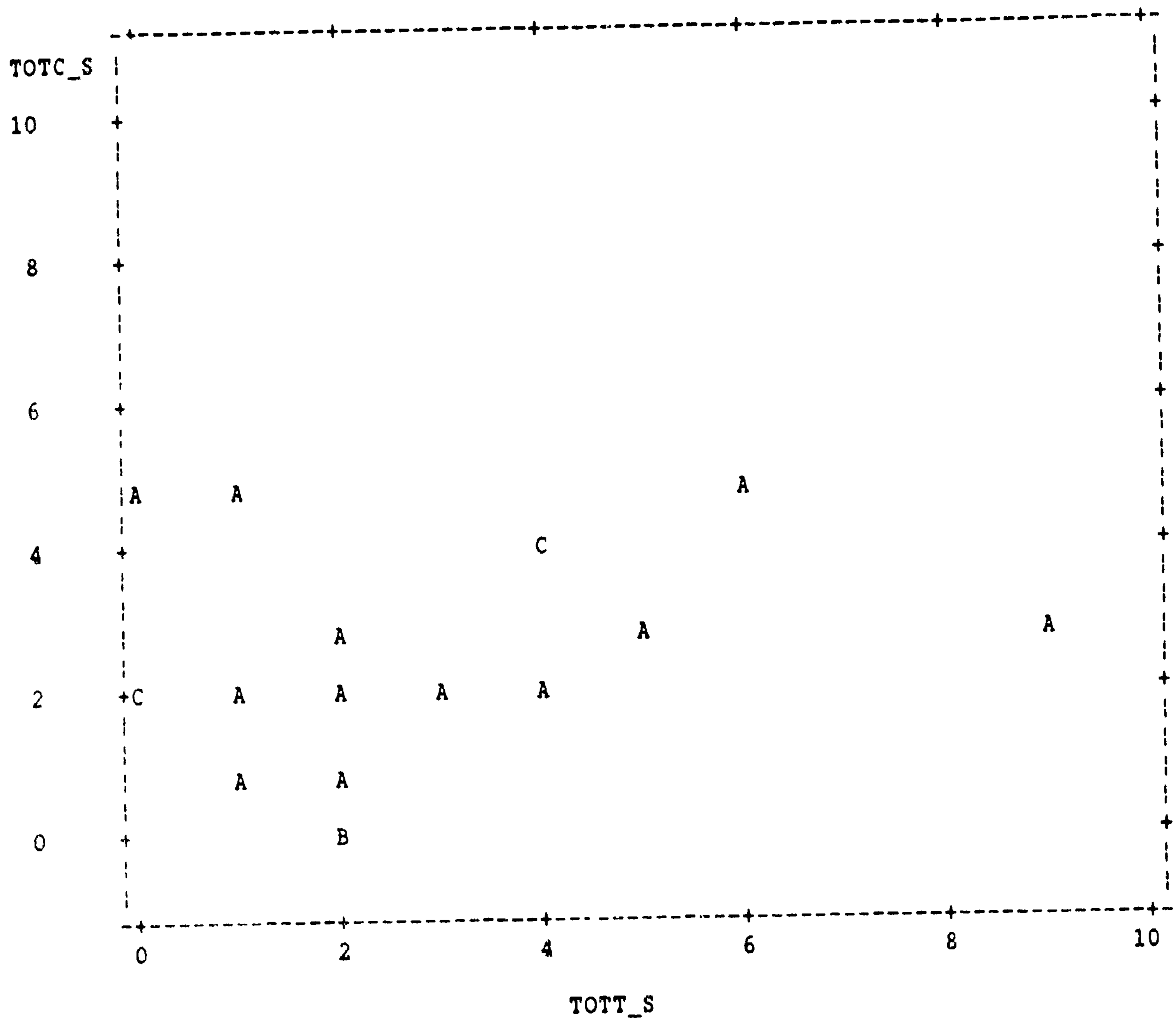


NOTE: 1 obs had missing value or was out of range

KEY: TOTC_A Addition error detected by computer
             TOTT_A Addition error detected by tester

REMEDIAL TEACHER

Plot of TOTC_S*TOTT_S   Legend: A = 1 obs, B = 2 obs, etc.



KEY: TOTC_S Subtraction error detected by computer
        TOTT_S Subtraction error detected by tester

CLASSROOM TEACHER

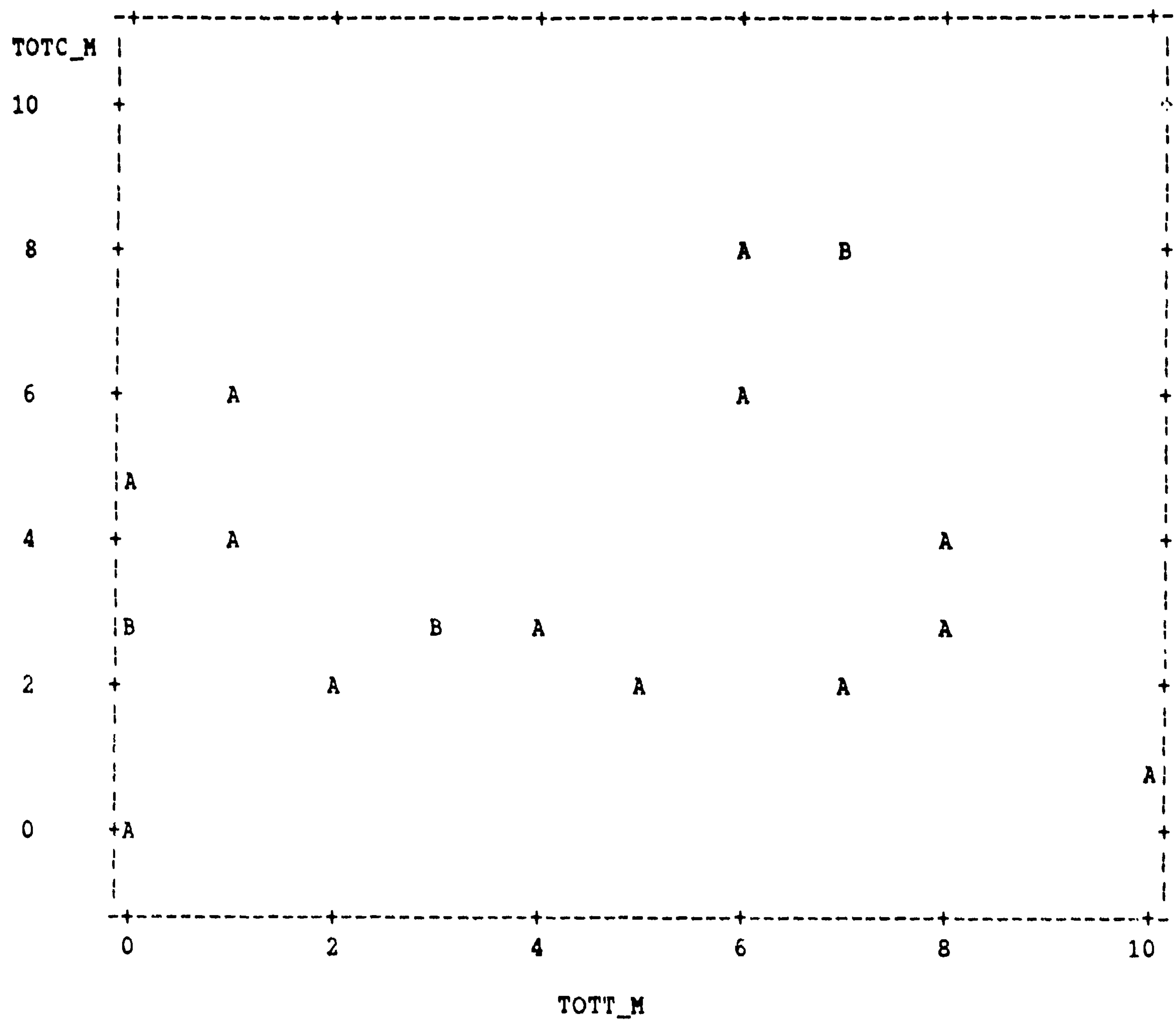
```
              Plot of TOTC_S*TOTT_S    Legend:  A = 1 obs, B = 2 obs, etc.

        --+-----------+-----------+-----------+-----------+-----------+-
 TOTC_S   |                                                             |
          |                                                             |
 10       +                                                             +
          |                                                             |
          |                                                             |
          |                                                             |
          |                                                             |
  8       +                                                             +
          |                                                             |
          |                                                             |
          |                                                             |
          |                                                             |
  6       +                                                             +
          |                                                             |
          |                                                             |
          |                                                             |
          |                                                             |
  4       +A     A                 A                                    +
          |                                                             |
          |                                                             |
          |                        A     A                              |
          |                                                             |
  2       +B     A     C     A     A                                    +
          |                                                             |
          |                                                             |
          |B           A           A                                    |
          |                                                             |
  0       +A     A                                                      +
          |                                                             |
          |                                                             |
        --+-----------+-----------+-----------+-----------+-----------+-
          0           2           4           6           8          10

                                TOTT_S
```

NOTE: 1 obs had missing value or was out of range

KEY: TOTC_S Subtraction error detected by computer
            TOTT_S Subtraction error detected by tester

REMEDIAL TEACHER

Plot of TOTC_M*TOTT_M   Legend: A = 1 obs, B = 2 obs, etc.

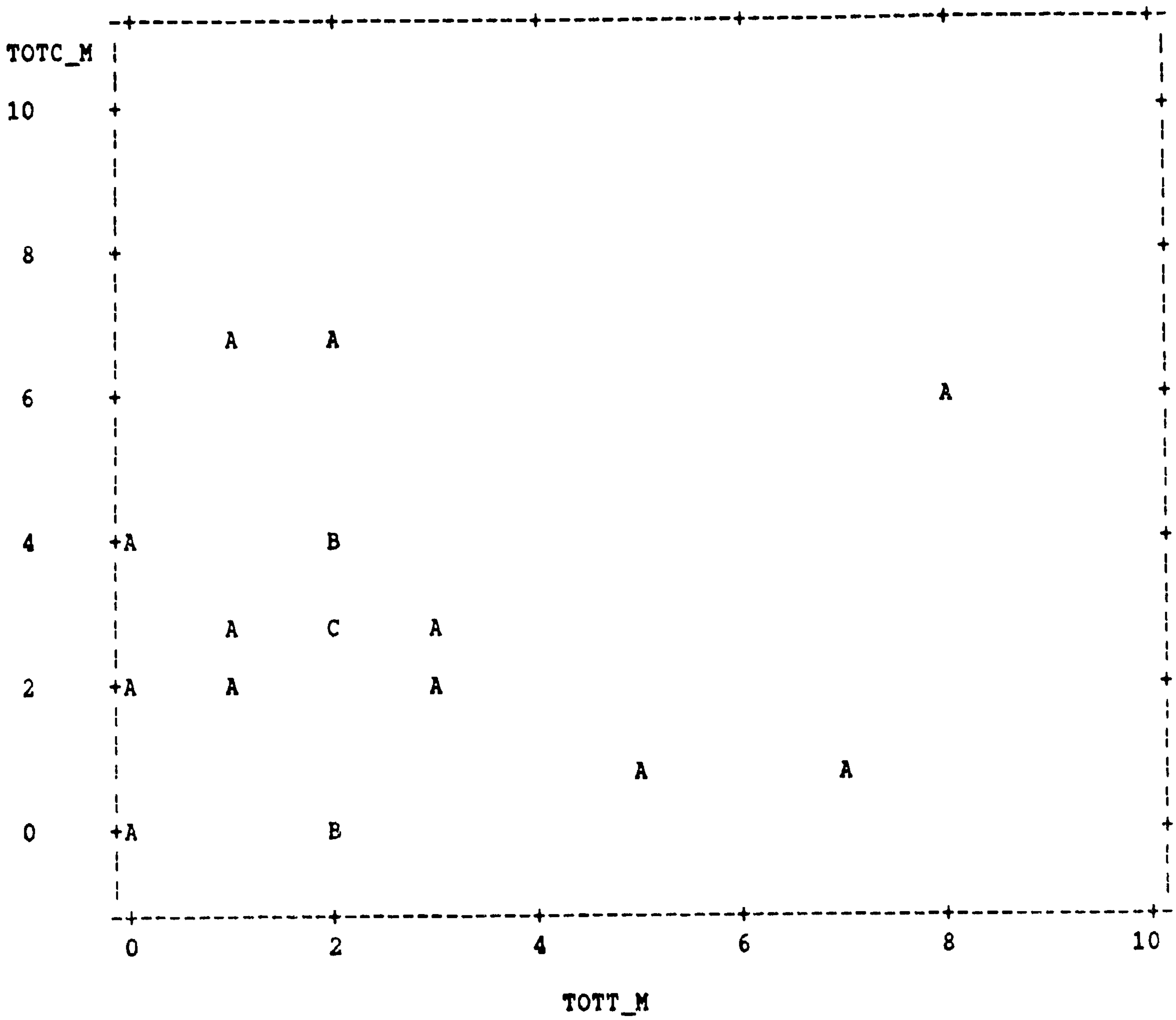


NOTE: 1 obs had missing value or was out of range

KEY: TOTC_M Multiplication error detected by computer
         TOTT_M Multiplication error detected by tester

CLASSROOM TEACHER

Plot of TOTC_M*TOTT_M   Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 1 obs had missing value or was out of range

KEY: TOTC_M Multiplication error detected by computer
TOTT_M Multiplication error detected by tester

APPENDIX C

NUMBER OF LEVELS IN THE SEVILLE DIAGNOSTIC ARITHMETIC TESTS IN WHICH ERRORS DETECTED BY COMPUTER MATCHED THE NUMBER DETECTED BY TESTERS

1986 RESULTS

ADDITION SUBTEST

PUPIL NUMBER	COMPUTER MORE	EXACT MATCH	COMPUTER FEWER	LEVEL TOTAL
1	3	10	0	13
2	0	9	0	9
3	1	7	1	9
4	1	10	3	14
5	0	8	1	9
6	1	11	2	14
7	1	7	1	9
8	0	9	0	9
9	1	8	0	9
10	1	8	0	9
11	1	8	0	9
12	1	8	0	9
13	.	.	.	.
14	0	6	1	7
15	3	10	0	13
16	3	6	1	10
17	0	8	1	9
18	2	7	0	9
19	1	7	1	9
20	1	6	1	8
21	0	9	0	9
22	3	7	3	13
23	3	7	0	10
24	1	2	0	3
25	1	1	0	2
26	3	5	2	10
27	2	11	4	17
28	1	10	0	11
29	3	6	0	9
30	1	8	0	9
31	3	5	1	9
32	2	7	0	9
33	0	9	0	9
34	1	6	2	9
35	2	9	2	13
36	2	7	0	9
37	2	3	0	5
38	1	7	2	10
39	2	7	1	10
40	1	7	1	9

SUBTRACTION SUBTEST

PUPIL NUMBER	COMPUTER MORE	EXACT MATCH	COMPUTER FEWER	LEVEL TOTAL
1	1	6	2	9
2	4	6	0	10
3	1	4	0	5
4	1	6	0	7
5	1	6	0	7
6	1	4	3	8
7	0	6	1	7
8	1	6	0	7
9	1	2	2	5
10	0	5	0	5
11	1	5	1	7
12	0	7	1	8
13	1	4	0	5
14	4	3	1	8
15	0	5	4	9
16	0	6	2	8
17	0	6	0	6
18	0	6	1	7
19	2	6	0	8
20	1	0	1	2
21	0	3	1	4
22	0	2	1	3
23	0	3	0	3
24	0	3	0	3
25	0	3	0	3
26	2	1	0	3
27	0	3	0	3
28	1	3	1	5
29	0	3	1	4
30	0	3	0	3
31	0	4	0	4
32	0	4	0	4
33	0	7	1	8
34	0	4	0	4
35	1	6	0	7
36	1	5	0	6
37	1	3	0	4
38	0	3	2	5
39	0	4	1	5
40	1	2	5	8

MULTIPLICATION SUBTEST

PUPIL NUMBER	COMPUTER MORE	EXACT MATCH	COMPUTER FEWER	LEVEL TOTAL
1	4	4	1	9
2	0	6	0	6
3	0	5	2	7
4	3	2	1	6
5	2	4	0	6
6	1	5	0	6
7	0	4	3	7
8	0	9	0	9
9	1	10	2	13
10	2	7	1	10
11	2	9	2	13
12	1	11	1	13
13	0	8	1	9
14	3	1	0	4
15	0	8	2	10
16	2	4	1	7
17	0	7	0	7
18	4	6	0	10
19	1	1	5	7
20	2	2	0	4
21	0	1	3	4
22	2	1	0	3
23	0	1	3	4
24	1	1	0	2
25	1	1	1	3
26	0	3	1	4
27	4	1	0	5
28	1	7	0	8
29	0	2	2	4
30	.	.	.	.
31	1	1	3	5
32	2	1	0	3
33	1	3	0	4
34	4	5	2	11
35	1	1	5	7
36	0	7	0	7
37	0	6	1	7
38	1	5	1	7
39	.	.	.	.
40	1	3	2	6